Final report M3-internship Technical Medicine

The value of ultrasound in diagnosing giant cell arteritis: comparing the intima-media thickness of the temporal and axillary arteries between manual longitudinal, semi-automatic longitudinal and transversally compressed measurements

Student:
Alexander Lau (s1134965)

August 26, 2019

Technical Supervisor
Prof. Dr. W. Steenbergen

Medical Supervisors
Dr. C. Alves
Dr. H.J. Bernelot Moens

Process Supervisor UT
Drs. P.A. van Katwijk

Abstract

Background: Giant cell arteritis (GCA) is a common medium-/large-sized vessel vasculitis that can result in blindness. An accurate early GCA diagnosis can minimize such complications, but temporal artery biopsy (TAB), the diagnostic gold standard, is moderately sensitive (39%) and not timely. Ultrasound (US) is a promising alternative that can show GCA lesions, such as an enlarged arterial intima-media thickness (IMT). However, US is highly operator-dependent.

Objective: (1) Investigate the diagnostic value of US for GCA, compared to TAB; (2) investigate and possibly improve US reliability with a semi-automatic algorithm or eminence-based transversal compression technique.

Methods: A retrospective cohort and pilot study were conducted to reach objective 1 and 2, respectively. The retrospective cohort study included suspected GCA patients, whose initial visit to the Rheumatology Department of the Dutch hospital ZGT, was between 01-01-2017 and 06-01-2018. The pilot study entailed ten healthy subjects, who were examined by two sonographers and whose longitudinal images and transversal images with vessel compression were recorded for the common temporal artery (CTA), frontal temporal artery (FTA) and axillary artery (AX). IMT measurements were performed by both sonographers and twice with the semi-automatic algorithm. The manual longitudinal, semi-automatic longitudinal and transversally compressed IMT measurements were compared using the intraclass correlation coefficient (ICC).

Results: Retrospective cohort study: the sample encompassed 103 patients, 34 GCA and 69 non-GCA. 100 US and 46 TAB examinations were performed. US and TAB respectively had a 75.8% and 80.8% sensitivity, and both had a 100% specificity. Pilot study: the ICC between semi-automatic and manual longitudinal measurements was moderate for the CTA (0.75) and FTA (0.56), but good for the AX (0.81). The ICC between transversally compressed and manual longitudinal measurements was poor for the CTA (0.27), FTA (0.25) and AX (0.14). The ICC between the manual longitudinal measurements was moderate for the CTA (0.57) and FTA (0.60), but good for the AX (0.87). The ICC between the transversally compressed measurements was moderate for the CTA (0.52), FTA (0.63) and AX (0.71). Lastly, the ICC between the semi-automatic longitudinal measurements was very good for the CTA (1.00), FTA (0.99) and AX (1.00).

Conclusion: US is a viable alternative to TAB for diagnosing GCA early. Semi-automatic longitudinal IMT measurements are as valid as manual longitudinal measurements and highly reliable. Transversally compressed IMT measurements are less valid than manual longitudinal measurements, but the reliability is comparable for the CTA and FTA, and lower for the AX.

Table of contents

List of abbreviations

95% CI	95% confidence interval
AX	Axillary artery
CRP	C-reactive protein
CTA	Common temporal artery
DICOM	Digital Imaging and Communications in Medicine
ESR	Erythrocyte sedimentation rate
EULAR	European League Against Rheumatism
FTA	Frontal temporal artery
GCA	Giant cell arteritis
IMT	Intima-media thickness
IQR	Interquartile range
NPV	Negative predictive value
PPV	Positive predictive value
PTA	Parietal temporal artery
ROI	Region of interest
SD	Standard deviation
TA	Temporal artery
TAB	Temporal artery biopsy
US	Ultrasound

1 Introduction

1.1 Clinical background

Giant cell arteritis (GCA) is a common medium-/large-sized vessel vasculitis that can result in blindness (1–3). GCA afflicts women 2-3 times as often as men and almost exclusively people aged over 50 years, with an incidence of 17-23 per 100,000 (4–6). Temporal headache and jaw claudication are the most recognizable symptoms, but systemic symptoms (e.g. fever) can also occur. (1,7–10) GCA complications are often irreversible, necessitating early diagnosis and corticosteroid treatment (1). Diagnosing GCA early is difficult due to its low incidence, generic symptoms and indistinct inflammation markers (e.g. erythrocyte sedimentation rate (ESR)) (11,12). (13) Moreover, the gold standard for diagnosing GCA is temporal artery biopsy (TAB), which has a low sensitivity (39%) and can take 1-2 weeks for the result to be known. (2) Early corticosteroid treatment decreases the risk of GCA complications, but increases the risk of corticosteroid-related complications (e.g. osteoporosis) (3,14). It can also cause subsequent imaging and histology results to be false-negative due to its anti-inflammatory effect (2,15,16). An accurate early diagnosis of GCA is necessary to minimize both GCA and corticosteroid complications.

1.2 Technical background of ultrasound

Ultrasound (US) may be a promising alternative to TAB for diagnosing GCA early. (12) US is inexpensive, real-time, noninvasive and does not use radiation or contrast agents (17). US is based on the emission and reception of sound waves reflected from tissue structures. This is made possible by piezoelectric crystals in the US probe, which convert alternating current into sound waves and vice versa. The most ubiquitously used US mode is Brightness (B-)mode, which is the translation of received sound waves into a grayscale image. A transmission of US contains a bandwidth of frequencies. The greater this frequency bandwidth, the higher the resolution in the US image. A frequency of at least 15 MHz is recommended for imaging temporal arteries (TAs) (18), because a resolution of at least 0.01 mm is required to clinically assess the arterial wall. The timing and intensity of the reflected sound wave determine the pixel location and brightness in the US image. Locating and assessing arteries for diagnosing GCA can be aided by integrating Doppler signals in US images. The Doppler effect is the frequency shift in relation to an observer that is moving relative to the emitting source. This frequency shift enables US to show the presence and direction of blood flow, and can assist in assessing the size of the lumen and the thickness of the arterial wall. (19–21)

1.3 US and the diagnosis of GCA

US examination of GCA conventionally includes the TA and axillary artery (AX), because both are easily accessible, with the former being one of the most frequently affected cranial arteries and the latter one of the most frequently affected extracranial arteries (11,12,22). The TA consists of three branches (i.e. common (CTA), frontal (FTA) and parietal (PTA)). (23) Other superficial arteries (e.g. carotid artery) can also be examined. (18) GCA causes inflammation that induces edema formation, hyperplasia of the intima and fibrosis of the arterial wall. (23–25) Indicators of GCA are a positive halo, compression sign (see Figure 1), and an abnormally thick intima-media thickness (IMT). The halo sign is positive if a hypoechoic swelling is visible of the arterial wall (26). If compression with the probe cannot make this swelling disappear, then the compression sign is also positive (27). (2,12,28) The arterial wall consists of three layers, from proximal to distal with respect to the lumen: the intima, media and adventitia. The intima and adventitia are hyperechoic, and the media is hypoechoic. Conventionally, IMT measurements in longitudinal images are performed on the far wall (i.e. the vessel wall more distal to the surface), because the layer borders are more distinct than in the near wall (i.e. the

vessel wall more proximal to the surface) (29,30). Figure 2 shows the IMT measurement of the AX in a longitudinal image. Studies have found US to be a viable alternative to TAB in diagnosing GCA early (see Table 2).

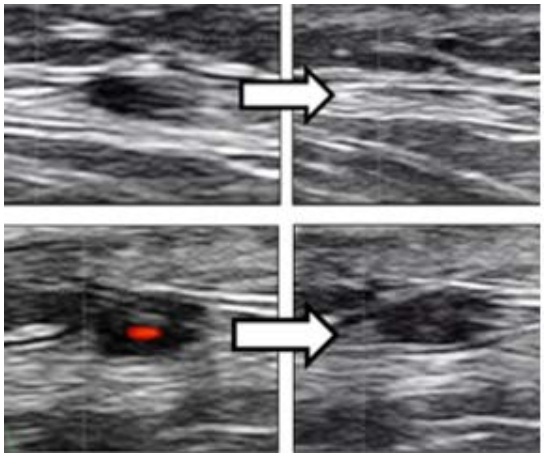


Figure 1: The halo sign and compression sign in transversal ultrasound (US) images of the temporal artery (TA) in giant cell arteritis (GCA). The upper images show a normal TA (left image) that is compressible (right image). The bottom two images show an abnormal TA with a positive halo sign (left image) and a positive compression sign (right image). (23) Schmidt WA. Ultrasound in the diagnosis and management of giant cell arteritis. Rheumatology (United Kingdom). 2018

Table 1: Thresholds of the intima-media thickness (IMT) for GCA in the TA branches (common (CTA), frontal (FTA) and parietal (PTA)) (31) and axillary artery (AX) (22,31) in US images.

Clinical assessment	CTA	FTA	PTA	AX
Normal	0.23 mm	0.19 mm	0.20 mm	0.6 mm
Likely pathological	≥ 0.42 mm	≥ 0.34 mm	≥ 0.29 mm	≥ 1.0 mm
Pathological	≥ 0.65 mm	≥ 0.54 mm	≥ 0.50 mm	≥ 1.5 mm

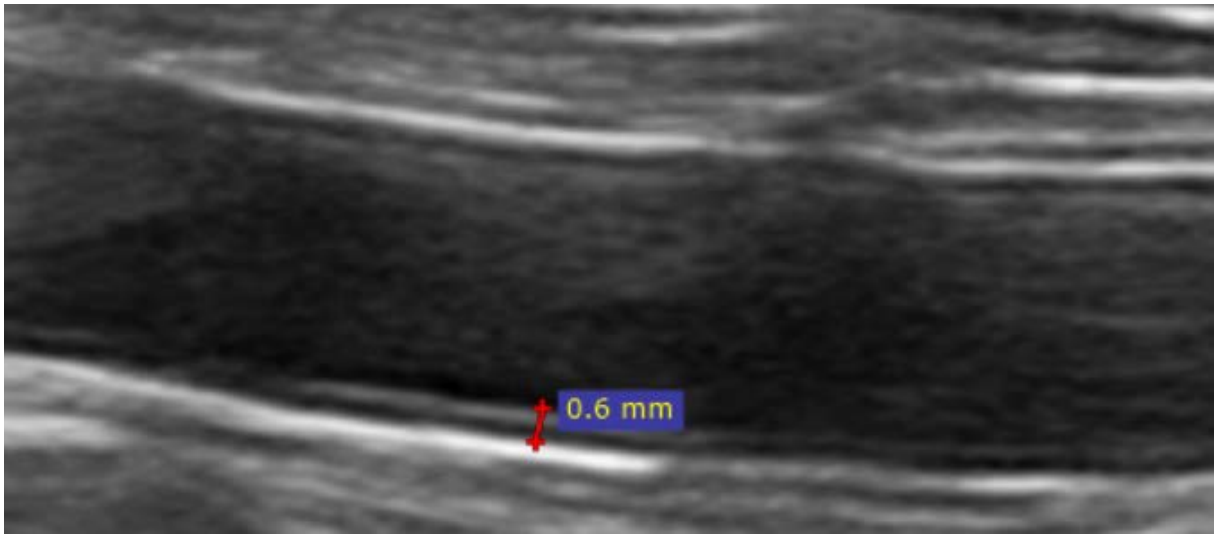


Figure 2: Measurement of the IMT of the AX far wall (i.e. more distal to the surface) in a longitudinal image. The arterial wall consists of three layers, from proximal to distal with respect to the lumen: the intima, media and adventitia. The intima and adventitia are hyperechoic, and the media is hypoechoic. The IMT is measured as 0.6 mm.

Table 2: Literature overview of the value of US in diagnosing and assessing GCA

First author and date	Title	Method	Results	Conclusion
Luqmani et al. (2016) (2)	The Role of Ultrasound Compared to Biopsy of Temporal Arteries in the Diagnosis and Treatment of Giant Cell Arteritis (TABUL): a diagnostic accuracy and cost-effectiveness study	A prospective multicenter cohort study was conducted that compared the sensitivity, specificity, interrater reliability and monetary cost of US and TAB in diagnosing GCA. The sample consisted of 381 newly suspected GCA patients, who underwent both US of the TAs and AXs, and TAB within seven days of starting treatment of high-dose corticosteroids. 257 patients were actually diagnosed with GCA in a follow-up assessment within six months.	The sensitivity and specificity for TAB were 39% and 100%, respectively. US had a higher sensitivity of 54%, but lower specificity of 81%. If TAB were only performed on patients with a negative US result, then the sensitivity and specificity of this strategy would be 65% and 81%, respectively. This would reduce the need for TAB by 43%. Combining the physician's assessment with TAB had a sensitivity and specificity of 91% and 81%, respectively, and with US a sensitivity and specificity of 93% and 77%, respectively. The combination with US would save £485 per patient compared to TAB.	US has a higher sensitivity, but lower specificity than TAB in diagnosing GCA. US is also more cost-effective and may reduce the diagnostic need for TAB.
Monti et al. (2018) (11)	The proposed role of ultrasound in the management of giant cell arteritis in routine clinical practice	A retrospective study was conducted to investigate the value of US in diagnosing and monitoring GCA. The sample consisted of 293 patients with suspected GCA, for 118 of whom an expert rheumatologist confirmed the diagnosis GCA. Each patient underwent US of the TAs and AXs. A positive halo sign and compression sign were considered indicative of GCA. US was also performed at each visit of follow-up. If patients were suspected of having flares, they were required to have additional assessment visits. Flares were defined as new or returning GCA symptoms and/or an unexplained increase of inflammation markers.	The sensitivity and specificity for US were 63.3% and 100%, respectively. The positive and negative predictive values (PPV and NPV) were 100% and 64.5%, respectively. Sensitivity was 81.8% for patients with jaw claudication and high inflammation markers. During follow-up, 21% of asymptomatic patients and 37% of flare patients had a positive US result, respectively. The number of halos reduced during follow-up. Only new and flaring patients had halos in four or more sites. The AX halo diameter decreased from 1.6 mm at initial presentation to 1.4 mm during follow-up.	US has a high PPV for diagnosing GCA. US can also be used to follow-up GCA patients by monitoring the number of halos and the AX halo diameter.
De Miguel et al. (2013) (32)	The utility and sensitivity of colour Doppler ultrasound in monitoring changes in giant cell arteritis	A blind prospective study was conducted to investigate the ability of US to detect change in GCA. The sample consisted of 30 GCA patients. 38 cases of GCA were included, 25 of which were first episodes and 13 were relapses. Each patient's baseline US showed at least one affected branch of the TA. Follow-up US was	Follow-up US showed that halos disappeared in 95% of cases, after a mean time of 11 weeks. For 50% of cases the halo sign became negative within 8 weeks. The results also showed that cases of first episode GCA were associated with a higher erythrocyte sedimentation rate (ESR), more affected TA	The halo sign often remains positive during the first two months after the start of corticosteroid therapy. US is a viable imaging modality in monitoring GCA activity.

		performed every two weeks during the first month and every subsequent four weeks, until no TA halo was visible.	branches and a longer time for halo disappearance, than cases of relapse. Lastly, patients with a higher number of affected TA branches needed more time for halo disappearance.	
Schäfer et al. (2017) (31)	Ultrasound cut-off values for intima-media thickness of temporal, facial and axillary arteries in giant cell arteritis	A prospective study was conducted to establish threshold values for arterial IMTs in order to distinguish between vasculitic and non-pathological arteries in suspected GCA. The sample entailed 80 patients, split into two groups, equal in size (i.e. 40 patients each) and matched in age and sex. One group consisted of newly diagnosed GCA patients, while the other served as the control group. IMT measurements were performed with US within 24 hours after the diagnosis. Each patient had the following arteries scanned bilaterally: the CTA, FTA, PTA, AX and facial artery.	The mean IMT of the control group for the CTA, FTA, PTA, facial artery and AX was 0.23, 0.19, 0.20, 0.24 and 0.59, respectively. The mean IMT of the GCA group was respectively 0.65, 0.54, 0.50, 0.53 and 1.7 mm for the CTA, FTA, PTA, facial artery and AX. The threshold IMT between the control and GCA group for the CTA, FTA and PTA, facial artery and AX was 0.42, 0.34, 0.29, 0.37 and 1.0 mm, respectively. The sensitivity and specificity were 100% for the CTA, FTA and AX. For the PTA and facial artery, the sensitivity was 97.2% and 87.5%, and the specificity was 98.7% and 98.8%, respectively. The intraclass correlation coefficient (ICC) for the interrater reliability was between 0.87 and 0.98.	IMT measurements with US can accurately differentiate vasculitic and non-pathological arteries in suspected GCA patients.
Diamantopoulos et al. (2014) (33)	Diagnostic Value of Color Doppler Ultrasonography of Temporal Arteries and Large Vessels in Giant Cell Arteritis: A Consecutive Case Series	A retrospective study was conducted to investigate the value of US in diagnosing GCA by scanning the TAs, AXs and common carotid arteries. The sample encompassed 88 patients with suspected GCA, of whom 46 patients were diagnosed with GCA. Every patient underwent US of the TAs, AXs and common carotid arteries. A positive halo sign was seen as indicative of GCA. 58 patients also underwent TAB within ten days of corticosteroid therapy, of whom 39 were diagnosed GCA patients.	The sensitivity and specificity of only scanning the TAs was 96% and 90%, respectively. When including the scans of the AXs, these percentages rose to 98% and 91%, respectively. Scanning the TAs, AXs and common carotid arteries resulted in a sensitivity and specificity of 100% and 95%, respectively. Lastly, the sensitivity and specificity of TAB were 67% and 95%, respectively.	US is a viable alternative to TAB in diagnosing GCA patients.

An important drawback of US is its high dependence on the sonographer's expertise. Luqmani et al. conducted a prospective multicenter cohort study with n=381 newly suspected GCA patients and compared US with TAB in diagnosing GCA. US had a higher sensitivity (54% versus 39%), but lower specificity (81% versus 100%) than TAB. The moderate sensitivity of US may be explained by its moderate interrater reliability (intraclass correlation coefficient (ICC): 0.61). 20 of the 24 sonographers were inexperienced in examining GCA. This lack of experience may have led to significant discrepancies in interpreting the acquired images (e.g. measuring the IMT in the image) and in conducting the US examination (e.g. applying pressure with the probe). Luqmani et al. concluded that discrepancies in US image interpretation and examination both contributed equally to the moderate interrater reliability. (2)

The reliability of image interpretation may be increased by partially automating IMT measurements in acquired images. Various fully automatic algorithms have been used to measure the IMT of carotid arteries with encouraging results (17,34–38). However, these fully automatic algorithms can often only be performed on standard arterial images with distinct arterial wall layers. US images of arteries are more heterogeneous in clinical practice, especially when examining TAs, which often only comprise a small part of the entire image. A semi-automatic algorithm that allows the sonographer to guide the automated IMT measurements (e.g. manual region of interest (ROI) selection), may be more suitable in clinical practice to improve the reliability of image interpretation.

An eminence-based transversal compression technique may reduce the measurement variability in transversal images, thereby increasing the reliability of US examination. IMT measurements are conventionally more performed in longitudinal than transversal images, but the former are also more difficult to acquire when examining TAs. IMT measurements are prone to overestimation in transversal images, due to the oblique probe application to maintain the color Doppler signal, and due to the tendency of the signal to only represent the faster blood flow in the center of the lumen. The transversal compression technique may mitigate these issues and is performed by applying pressure with the probe on the artery obliquely until the Doppler signal disappears; by subsequently repositioning the probe perpendicularly to the artery (while maintaining pressure) and dividing the height of the compressed artery by two, the IMT may be measured more accurately than in a transversal image without vessel compression (see Figure 3). Improving the reliability of US image interpretation and examination, may also increase the sensitivity and specificity of US in diagnosing GCA.

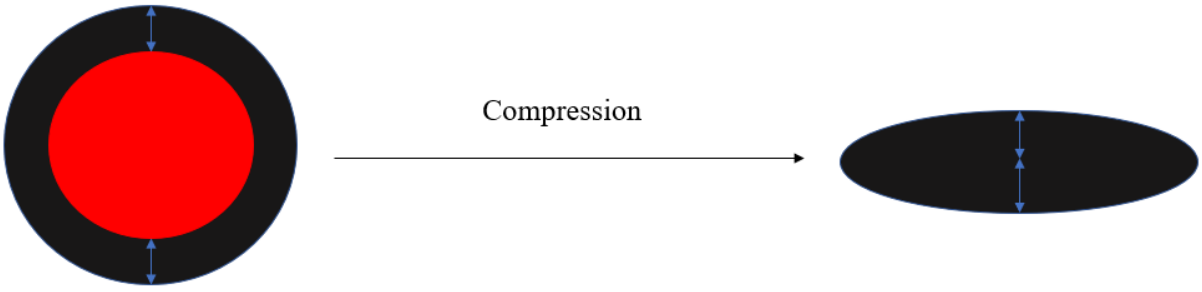


Figure 3: Transversal US measurement of the IMT with an eminence-based transversal compression technique. The intima-media is depicted as black, the Doppler signal of the lumen as red and the IMT with a blue double-headed arrow. The transversal compression technique is performed by applying pressure with the probe on the artery obliquely until the Doppler signal disappears; by subsequently repositioning the probe perpendicularly to the artery (while maintaining pressure) and dividing the height of the compressed artery by two, the IMT may be measured more accurately than in a transversal image without arterial compression.

1.4 Research goal and objectives

The primary research goal is to improve the reliability of US in measuring the IMT in order to possibly increase the value of US in diagnosing GCA early. The current study was conducted in a Dutch hospital. Although studies (see Table 2) have shown US to be diagnostically valuable, no such study has been conducted in the Netherlands. The current study can therefore provide additional insights in the external validity of previous studies (39). The following objectives have been set towards the primary research goal:

1. Investigate the value of US of the TAs and AXs in diagnosing GCA early in suspected patients, compared with TAB.
2. Improve the reliability of US of the TAs and AXs in measuring the IMT in healthy subjects by:
    a. developing a semi-automatic algorithm to measure the IMT in longitudinal images and comparing these measurements with the manual IMT measurements in longitudinal images.
    b. comparing manual IMT measurements in transversal images with compression and longitudinal images.

A retrospective cohort study and pilot study were conducted to reach the objectives. Objectives 1 and 2b were achieved with the retrospective cohort study and pilot study, respectively. Objective 2a was attained with both the retrospective cohort study and the pilot study. Figure 4 shows a research overview.

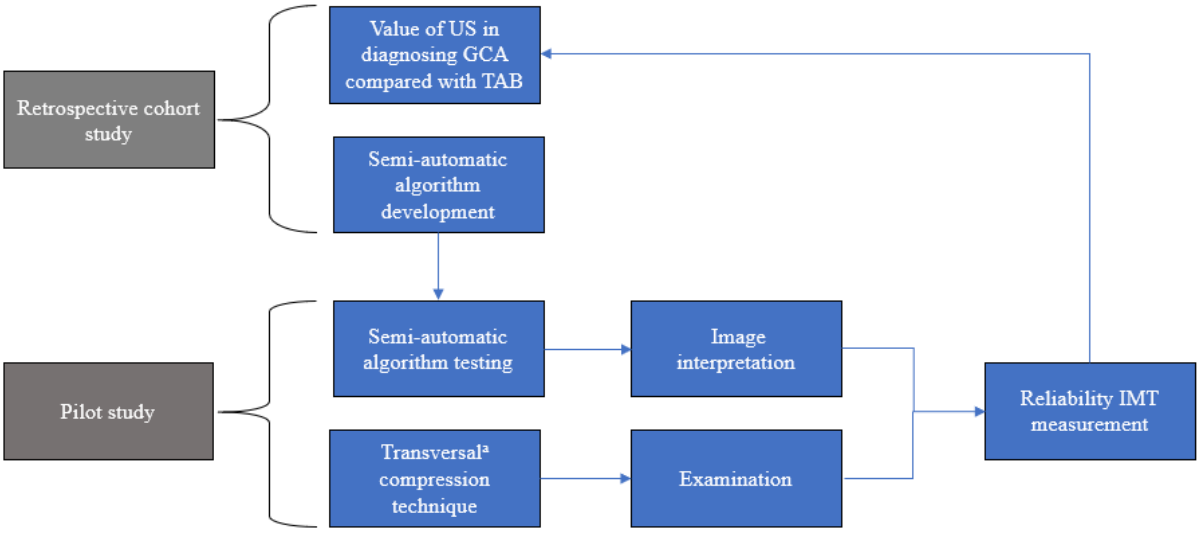


Figure 4: Research overview. A retrospective cohort study was conducted to investigate the value of US in diagnosing GCA compared with TAB, and to develop a semi-automatic algorithm to measure the IMT in longitudinal images. A pilot study was also conducted to investigate the validity and reliability of the semi-automatic algorithm and an eminence-based transversal compression technique in measuring the IMT manually in transversal images, by comparing them both with manual IMT measurements in longitudinal images. The semi-automatic algorithm and transversal compression technique may improve the interrater reliability of IMT measurements, which in turn may increase the value of US in diagnosing GCA. [a]Eminence based.

2 Methods

A retrospective cohort study and pilot study were conducted at the Rheumatology Department of the hospital ZGT in Almelo and Hengelo, the Netherlands, to achieve the two research objectives.

2.1 Retrospective cohort study

2.1.1 Research sample

The sample encompassed patients whose initial visit to the Department concerning a GCA suspicion was between January 1, 2017, and June 1, 2018. These patients were either referred to the Department by a general practitioner or medical specialist, or they were already under a rheumatologist's supervision at the Department for another condition (e.g. polymyalgia rheumatica). The reference diagnosis was defined as the documented GCA diagnosis by the treating physician, six months after the initial visit to the Department. The reference diagnosis was not solely based on the TAB, but on all the data available to the treating physician. Patients without a reference diagnosis were excluded. The referral for TAB was not based on a standard protocol, but on the treating physician's own assessment of the probability of GCA. This assessment for TAB referral could have included the US result, but was not necessarily based on it. After a patient was referred for TAB, the surgeon excised 1-2 cm of the TA and sent it to the pathologist. The TAB result was known after 7-10 days. The following demographic and clinical data were collected: age, gender, symptoms, C-reactive protein (CRP (mg/l)), ESR (mm/h), corticosteroid use and the US and TAB results, which were based on the assessment of the sonographer and pathologist, respectively. The documented diagnoses after three and 12 months were also collected to examine the robustness of the reference diagnosis.

2.1.2 US variables of GCA

The US examinations were part of the regular diagnostic process of GCA and conducted with an Esaote MyLab Twice eHD (Esaote, Genoa, Italy). The CTA, FTA and PTA were imaged with an 18 MHz or 22 MHz linear probe, and the AX with a 13 MHz linear probe. The pulse repetition frequencies for the TA branches and AX were 2-3 kHz and 3-3.5 kHz, respectively. The focus was 5 mm for the TA branches and variable for the AX. The image depth was 15 mm for the TA branches and variable for the AX. The TA branches and AXs were imaged bilaterally in the longitudinal and transverse planes. The halo sign, compression sign and IMT (mm) were noted during the original examination. The halo sign and compression sign were operationalized as positive if they were present either unilaterally or bilaterally in at least one TA branch or AX. The clinical assessment of the IMT of each individual artery ('normal', 'likely pathological' and 'pathological') was determined with the thresholds in Table 1 (31). The IMT was operationalized as 'normal' if bilaterally every IMT of the TA branches and AXs was assessed as such. The IMT was operationalized as 'likely pathological' if the IMT of at least one TA branch or AX was assessed as such, without a single TA branch or AX being assessed as 'pathological'. The IMT was operationalized as 'pathological' if the IMT of at least one TA branch or AX was assessed as such.

2.2 Pilot study

2.2.1 Research sample and examination protocol

The subjects of the research sample were recruited among hospital staff members and volunteers between April and June 2019. The sample was predetermined to consist of ten healthy subjects, five men and five women. Inclusion criteria were an age of 50 years or older and a written informed consent to participate in the study. Two experienced sonographers,

denoted as A and B, examined each subject's CTA, FTA and AX bilaterally, recording longitudinal and transversal images. The PTA was excluded, because the IMT was not measurable in healthy subjects. The transversal images were recorded with vessel compression. To accurately compare the semi-automatic longitudinal, transversally compressed and manual longitudinal IMT measurements, the measurements had to be performed at the same location on the subject. The first sonographer therefore marked the places of the probe on the subject's skin where he recorded the longitudinal images, so that the second sonographer could put the probe in the same place. The transversal images were acquired by rotating the probe ¼ turn at the arterial segment in the middle of the longitudinal image. After saving each image, the sonographer denoted the specific IMT location in the image to be measured (see Figure 5). The order in which the sonographers examined a subject was decided arbitrarily, with five subjects being examined by sonographer A first and the other five by sonographer B first. Both sonographers measured the IMT in each image independently after all subjects were examined. The semi-automatic algorithm was also applied to every image by a researcher. The IMT in each image was semi-automatically measured twice to test its reproducibility.

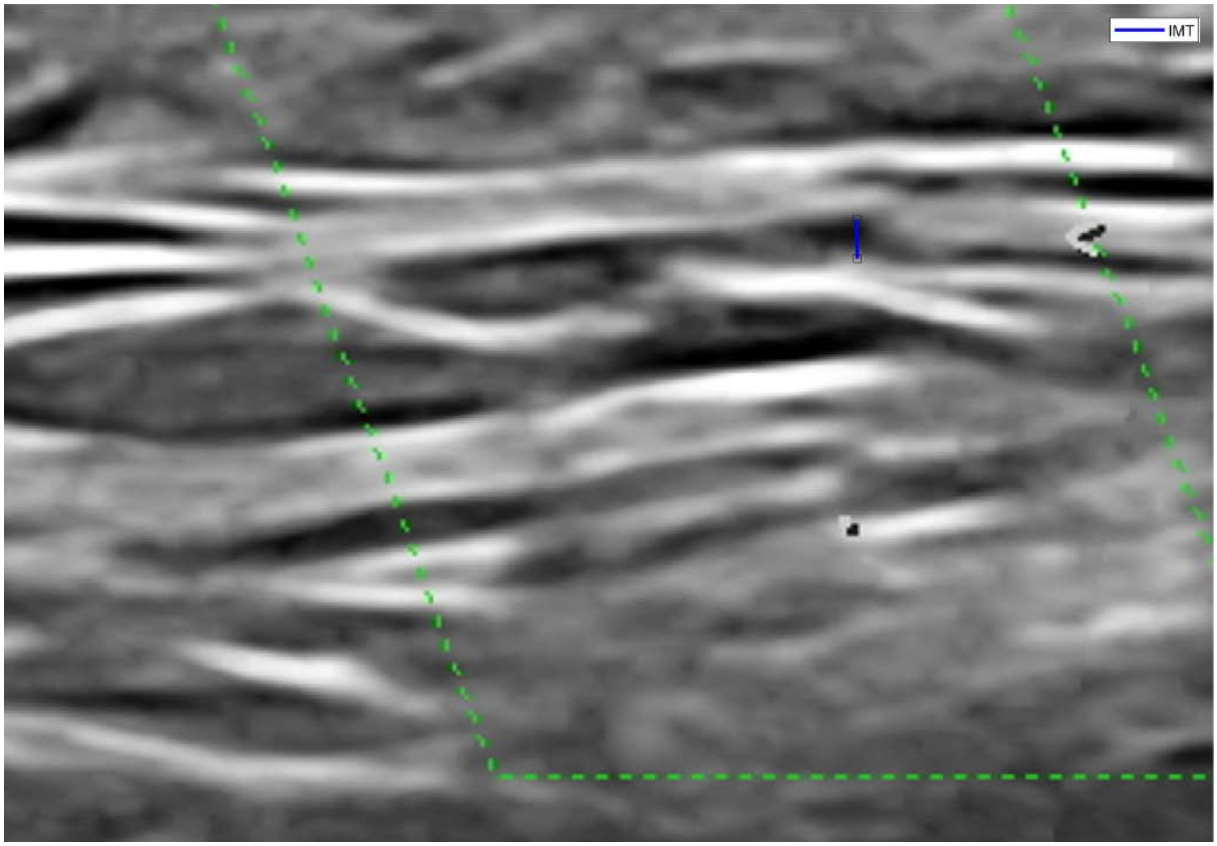


Figure 5: Transversal US image of the CTA with compression. The specific IMT location to be measured is denoted with a white dot and less-than sign. The dot denotes the horizontal location and the less-than sign the vertical location. The blue line is the sonographer's IMT measurement.

2.3 Statistics and materials

For the retrospective cohort study, The GCA and non-GCA patients were statistically compared on demographic and clinical characteristics. Categorical variables were compared with the chi-square test. Continuous variables were compared with the t-test. The Mann-Whitney U test was used for non-normally distributed continuous variables. The diagnostic value of US and TAB was operationalized with the sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). Their 95% confidence intervals (95% CIs) were calculated with an online calculator (40). The potential masking effect of corticosteroid use on US and

TAB results was also investigated. Corticosteroid use was defined as using corticosteroids up to six days before the examination (2,12).

Longitudinal TA branch and AX US images of patients in the retrospective cohort study were used to conduct preliminary IMT measurement tests to develop the semi-automatic algorithm. The maximum number of images included was predetermined to be 20. The inclusion criteria were that the image was stored on the US machines at the Rheumatology Department, and that the image had a visible arterial far wall with discernible intima, media and adventitia borders that were not obscured by a Doppler signal or manual IMT measurement signs. Images were excluded if they were similar to other images that were already included. This was judged qualitatively by a researcher who based his assessment on how similar the images looked in terms of the size, orientation and visibility of the arterial far wall and its layers. The results of the preliminary tests were evaluated qualitatively by the researcher, who visually inspected the IMT measurement in the image to see if it matched with the distance between the upper border of the intima and the upper border of the adventitia.

For the pilot study, the following IMT measurements were compared using the Bland-Altman plot (41–43) and the ICC with its 95% CI, based on a single-rating, absolute-agreement, two-way mixed-effects model (44):

1. Semi-automatic versus manual measurements in longitudinal images
    a. Semi-automatic measurements 1 versus sonographer A's measurements
    b. Semi-automatic measurements 2 versus sonographer A's measurements
    c. Semi-automatic measurements 1 versus sonographer B's measurements
    d. Semi-automatic measurements 2 versus sonographer B's measurements
2. Manual measurements in transversal images with compression versus manual measurements in longitudinal images
    a. Sonographer A's transversally compressed measurements versus sonographer A's longitudinal measurements
    b. Sonographer B's transversally compressed measurements versus sonographer B's longitudinal measurements

The next IMT measurements were only compared with the ICC:

3. Semi-automatic measurements 1 versus semi-automatic measurements 2
4. Sonographer A's transversally compressed measurements versus sonographer B's transversally compressed measurements
5. Sonographer A's longitudinal measurements versus sonographer B's longitudinal measurements

Comparison 1 and 2 investigated the validity of the semi-automatic algorithm and the transversal compression technique, while comparison 3 and 4 investigated the reliability of the semi-automatic algorithm and transversal compression technique. Comparison 5 was used as a benchmark for the validity and reliability of the other comparisons. The following ICC thresholds were applied to assess the validity and reliability: very good (>0.9 - 1.0); good (>0.75 - 0.9), moderate (0.5 - 0.75); poor (<0.5) (44). The Bland-Altman plots show the differences between the semi-automatic longitudinal / transversally compressed and the manual longitudinal IMT measurements. The Bland-Altman plots also contain the measurement bias and the statistical and clinical limits of agreement. The measurement bias was operationalized with the mean difference between the semi-automatic longitudinal / transversally compressed and the manual longitudinal IMT measurements. The statistical limits of agreement were set to +/- 2 standard deviations (SDs) of the manual longitudinal IMT measurements. The clinical limits of agreement were based on the clinical IMT thresholds in Table 1, and set to half of the interval between a 'normal' IMT and a 'pathological' IMT. The clinical limits of agreement for

the IMT measurements in the CTA, FTA and AX were respectively: +/- 0.21 mm, +/- 0.175 mm and +/- 0.45 mm.

The US images for the development of the semi-automatic algorithm and the pilot study, consisted of 555x800 (rows x columns) pixels and were stored in Digital Imaging and Communications in Medicine (DICOM) files. The manual longitudinal measurements were conducted in a DICOM viewer, specifically designed for the pilot study. The DICOM viewer automated ancillary tasks associated with IMT measurements and anonymized the sonographer and subject's identity annotations that were visible in the image. The DICOM viewer presented each image consecutively to the sonographer, who measured the IMT with an adjustable line. The DICOM viewer also offered the option of skipping an image if the IMT could not be determined. Afterwards, all images with IMT measurements and all quantitative measurement values were systematically saved. The DICOM viewer and semi-automatic algorithm were developed in MATLAB 2018b (MathWorks, Inc., Natick, MA, USA). SPSS (version IBM SPSS Statistics 25, IBM Corp.) was used for the statistical tests. A p-value lower than 0.05 was considered significant. The retrospective cohort study and pilot study were conducted according to the Declaration of Helsinki.

3 Results

3.1 Retrospective cohort study

3.1.1 Research sample

The research sample initially consisted of 104 patients with suspected GCA. One patient without a reference diagnosis was excluded. The final sample had 103 patients, encompassing 65 (63.1%) women and 38 (36.9%) men. The mean age was 71.5 years (SD: 11.3 years). 34 (33.0%) patients had the reference diagnosis of GCA and 69 (67.0%) patients did not. No reference diagnosis was overturned after 12 months and each reference diagnosis was already established after three months, indicating that the reference diagnosis is robust. 100 US examinations were performed in the sample, compared with only 46 TAB examinations. No patient had more than one US or TAB examination.

3.1.2 Demographic and clinical characteristics

The comparison between characteristics of GCA and non-GCA patients is shown in Table 3. The group of GCA patients contained significantly more women than the group of non-GCA patients. The mean age of GCA patients was not significantly higher. Jaw claudication was significantly more widespread among GCA patients and they also had relatively more temporal headaches, but this difference was not significant. The physical examination results of the TA showed the following significant differences: GCA patients experienced more often localized pain, thickening and an impalpable pulse. GCA patients also had significantly higher median values for ESR and CRP. Lastly, GCA patients had significantly more positive results for US and TAB. In fact, no non-GCA patient received a positive US or TAB result.

Table 3: Comparison of demographic and clinical characteristics between GCA and non-GCA patients. The percentage of each patient group for each variable only took into account the non-missing values. The percentages of missing values were calculated for the entire GCA or non-GCA patient group (34 and 69, respectively). A p-value lower than 0.05 was considered significant. Abbreviations: CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; IQR, interquartile range; SD, standard deviation.

	GCA patients n = 34	**Non-GCA patients n = 69**	**p-value**
Demographic characteristics			
Female; n (%)	26 (76.5%)	39 (56.5%)	0.048

Age (years); mean ± SD	74.1 ± 10.0	70.2 ± 11.7	0.104
Characteristic GCA symptoms			
Temporal headache; n (%)	19 (55.9%)	26 (37.7%)	0.080
Jaw claudication; n (%) Missing (%)	 14 (42.4%) 1 (2.9%)	 3 (4.6%) 3 (4.3%)	 <0.001
Physical examination of the TA			
Pain[a]; n (%) Missing (%)	 13 (43.3%) 4 (11.8%)	 11 (17.2%) 5 (7.2%)	 0.007
Thickening[a]; n (%) Missing (%)	 16 (53.3%) 4 (11.8%)	 2 (3.1%) 5 (7.2%)	 <0.001
Impalpable pulse [a]; n (%) Missing (%)	 5 (16.7%) 4 (11.8%)	 1 (1.6%) 5 (7.2%)	 0.012
Inflammation markers			
ESR (mm/h); Median (IQR) Missing (%)	 77.0 (47.0– 100.0) 3 (8.8%)	 36.0 (18.0 – 81.0) 3 (4.3%)	 0.001
CRP (mg/L); Median (IQR) Missing (%)	 50.5 (29.8 – 93.5)	 18.0 (3.25 – 58.0) 1 (1.4%)	 <0.001
Imaging and histology			
Positive US result; n (%) Missing (%)	 25 (75.8%) 1 (2.9%)	 0 (0.0%) 2 (2.9%)	 <0.001
Positive TAB result; n (%) Missing (%)	 21 (80.8%) 8 (23.5%)	 0 (0.0%) 49 (71.0%)	 <0.001

[a]Unilateral or bilateral

3.1.3 US variables of GCA

The comparison of GCA characteristics in US images between GCA and non-GCA patients is shown in Table 4. GCA patients were significantly more associated with a positive halo sign, positive compression sign and a 'pathological' IMT than non-GCA patients. 27.3% of GCA patients had a 'normal' IMT, compared with 79.1% of non-GCA patients. Unexpectedly, 9.1% of GCA patients had a 'likely pathological' IMT, compared with 17.9% of non-GCA patients. Lastly, 63.6% of GCA patients had a 'pathological' IMT, compared with 3.0% of non-GCA patients.

Table 4: Comparison of GCA characteristics in US images between GCA and non-GCA patients. The percentages of the positive and negative US results for each variable only took into account the non-missing values. The percentages of missing values were calculated for all the positive or negative results (34 and 69, respectively).

US characteristics of GCA	GCA patients n = 34	Non-GCA patients n = 69	p-value
Positive halo sign[a]; n (%) Missing (%)	 25 (75.8%) 1 (2.9%)	 5 (7.5%) 2 (2.9%)	 <0.001
Positive compression sign[a]; n (%) Missing (%)	 23 (69.7%) 1 (2.9%)	 2 (3.0%) 2 (2.9%)	 <0.001
IMT[b]; n (%) 'Normal' 'Likely pathological' 'Pathological'	 9 (27.3%) 3 (9.1%) 21 (63.6%)	 53 (79.1%) 12 (17.9%) 2 (3.0%)	 <0.001

Missing (%)	1 (2.9%)	2 (2.9%)	

[a]The halo sign and compression sign were operationalized as positive if they were positive in at least one TA branch (CTA, FTA or PTA) or one AX; [b]the IMT was operationalized as 'normal' if bilaterally every IMT of the TA branches and AXs was assessed as such, using the thresholds in Table 1; the IMT was operationalized as 'likely pathological' if the IMT of at least one TA branch or AX was assessed as such, without a single TA branch or AX being assessed as 'pathological'; the IMT was operationalized as 'pathological' if the IMT of at least one TA branch or AX was assessed as such.

3.1.4 Diagnostic parameters of US and TAB

The comparison between US and TAB results in patients who received both examinations is shown in Table 5. Every TAB examination in the sample was preceded by a US examination. 85.7% and 92.0% of patients with a positive and negative TAB result, respectively, also had a positive and negative US result, respectively. The results of TAB and US did not match for 8.0% and 14.3% of patients with a negative and positive TAB result, respectively. Knowing that no non-GCA patient received a positive US result, theoretically, 43.5% of all patients who underwent TAB could have been exempted from it.

The diagnostic parameters of US and TAB are displayed in Table 6. As expected from the US and TAB results in Table 3, the specificity and PPV for both examinations were 100%. TAB had a higher sensitivity than US, but US had a better NPV. Table 7 shows the difference in corticosteroid use between patients with a positive result and those with a negative result for US and TAB. Patients with a positive US result did not use significantly less corticosteroids than those with a negative US result. Conversely, relatively more patients with a positive TAB result, used corticosteroids than those with a negative TAB result, but this difference was not significant.

Table 5: Comparison between US and TAB results in patients who received both.

	Positive TAB result n = 21	**Negative TAB result n = 25**
Positive US result	18 (85.7%)	2 (8.0%)
Negative US result	3 (14.3%)	23 (92.0%)

Table 6: Diagnostic parameters of US and TAB. Abbreviations: 95% CI, 95% confidence interval; NPV, negative predictive value; PPV, positive predictive value.

Diagnostic parameters	**US**	**TAB**
Sensitivity (95% CI)	75.8% (57.4-88.3%)	80.8% (60.0-92.7%)
Specificity (95% CI)	100% (93.2-100%)	100% (80.0-100%)
PPV (95% CI)	100% (83.4-100%)	100% (80.8-100%)
NPV (95% CI)	89.3% (79.5-95.0%)	80.0% (58.8-92.4%)

Table 7: Comparison in corticosteroid use between patients with a positive result and those with a negative result for US and TAB.

	Positive US result n = 25	**Negative US result n = 75**	**p-value**
Corticosteroid use US[a]; n (%)	8 (32.0%)	22 (29.3%)	0.801
	Positive TAB result n = 21	**Negative TAB result n = 25**	**p-value**
Corticosteroid use TAB[a]; n (%)	15 (71.4%)	11 (44.0%)	0.062

[a]Corticosteroid use was defined as using corticosteroids up to six days before the examination.

3.1.5 Semi-automatic algorithm

The preliminary IMT measurement tests for the development of the semi-automatic algorithm had a predetermined sample of 20 images. 93 longitudinal images of 20 patients, of whom 13 had GCA, were reviewed in total, of which 41 satisfied the inclusion criteria. 17 AX and 4 CTA images were excluded due to their similarity to already included images. The sample eventually encompassed eight CTA, four FTA, four PTA and four AX images. These images belonged to four GCA and five non-GCA patients. The fully developed semi-automatic algorithm was able to measure the IMT in all 20 images. An overview of the semi-automatic algorithm is shown in Figure 6, with the algorithmic process illustrated with an FTA example in Figure 7-13. The only manual component of the algorithm concerns the selection of the ROI in the longitudinal US image. The ROI contains the adventitia, media and intima of the arterial far wall (see Figure 7). Edge detection is performed on the ROI with the Canny algorithm, which results in a binary image of the ROI with only the edges visible (see Figure 8). Edges can be identified as areas in an image where the intensity gradient is locally highest. The Canny algorithm uses a double threshold to assess these intensity gradients. Intensity gradients greater than the upper threshold are strong edges and automatically included in the binary edge image, and intensity gradients lower than the lower threshold are not considered edges and excluded. The intensity gradients between the two thresholds are weak edges. A weak edge is only included in the binary edge image if it is directly connected to a strong edge. (45)

To reliably determine which respective edges, correspond to the upper adventitial and intimal borders requires the adventitia and intima to be identified independently from the Canny edge detection algorithm. Firstly, the edges in the original ROI image are sharpened with the unsharp masking technique, which sharpens edges in images by subtracting identical lower resolution images from them, so that only sharp edges remain (46). This is followed by the binarization of the image with Otsu's method, which uses the binarization threshold that minimizes the intraclass variance of pixel intensities at either side of the threshold in the sharpened image (47) (see Figure 9). The structure pertaining to the adventitia in the binary image can be identified by its large size and high mean intensity value of the corresponding pixels in the original ROI image (Figure 10). Each structure in the image is assigned a value by multiplying its number of pixels and its mean intensity value in the original ROI image. The structure with the highest value is identified as the adventitia. The upper border of this adventitial structure is then determined for the comparison with the edge image in Figure 8.

Both the adventitia and intima are horizontal linear structures. To identify the intima, the sharpened ROI image is processed with a motion filter, which mimics the linear motion of a camera (48). This filter superimposes straight lines on the image, along which the intensity values of pixels are averaged. The length of this line spans the width of the ROI image and the orientation of the line (relative to the horizontal axis of the image) is set equal to the orientation of the adventitial structure (relative to the horizontal axis of the image) previously determined in Figure 9. The filtered image is subsequently binarized with Otsu's method (see Figure 11) and the linear structure most closely above the adventitial structure is identified as the intima (see Figure 12). The upper border of this intimal structure is analogously determined for the comparison with the binary edge image in Figure 8. The application of a motion filter may seem superfluous, as the intimal structure could already be identified in Figure 9, but it is essential in images where the intima is poorly visible. Next, the adventitial and intimal structures are used to identify the upper borders of the adventitia and intima in the binary edge image (Figure 13). The adventitial and intimal structures could also be used to measure the IMT without the binary edge image, but their upper borders are less accurate due to the distortion of the motion filter.

The IMT is calculated per pixel column as the height difference between the two borders and multiplied with the scale of the image, in this example 0.308 mm/pixel. Lastly, the IMT is multiplied with the cosine of the orientation of the adventitia (relative to the horizontal axis of the image) to correct for the orientation of the arterial wall. The mean of the corrected IMT values is taken as the semi-automatic IMT measurement.

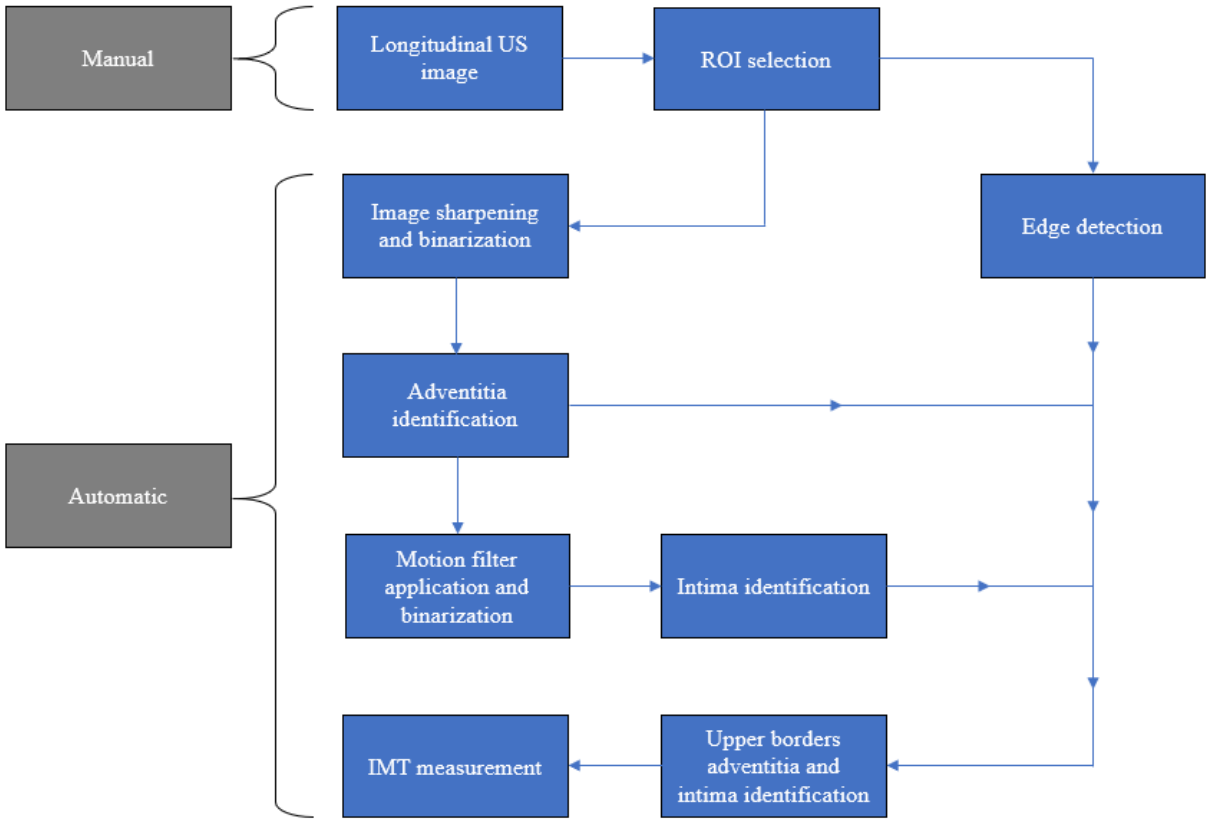


Figure 6: Overview of the semi-automatic algorithm for IMT measurements in longitudinal US images.

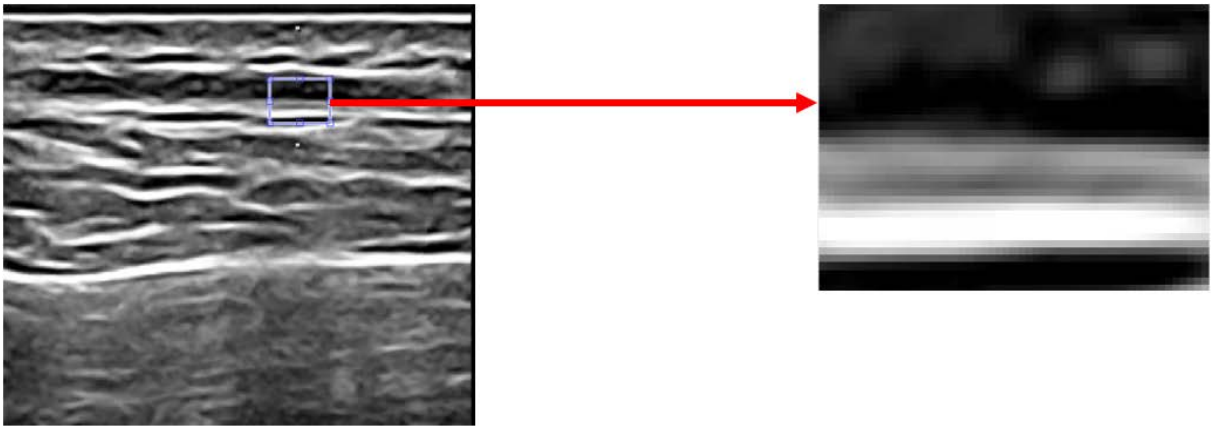


Figure 7: ROI selection. The original US image shows the FTA. The white dots in the image were placed by the sonographer during the examination to indicate the segment of the artery to be measured.

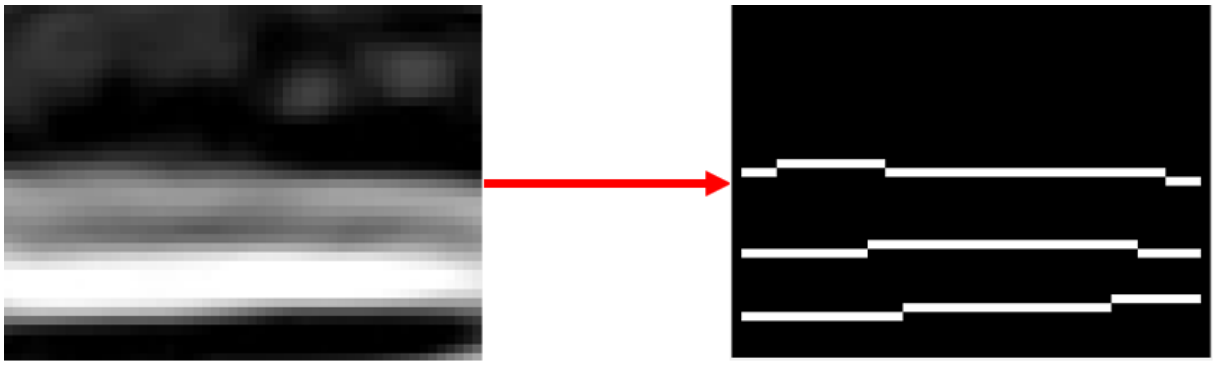


Figure 8: Edge detection. Edge detection is performed on the ROI (left image) with the Canny algorithm (45), which results in a binary image (right image) of the ROI with only the edges visible.

Figure 9: Image sharpening and binarization. The edges in the original ROI image (left image) are sharpened with the unsharp masking technique (46). The sharpened image (middle image) is then binarized with Otsu's method (47) (right image).

Figure 10: Adventitia identification. Each structure in the original binary image (left) is assigned a value by multiplying its number of pixels and its mean intensity value in the original ROI image. The structure with the highest value is identified as the adventitia (middle image). The upper border of this adventitial structure is then determined (right image).

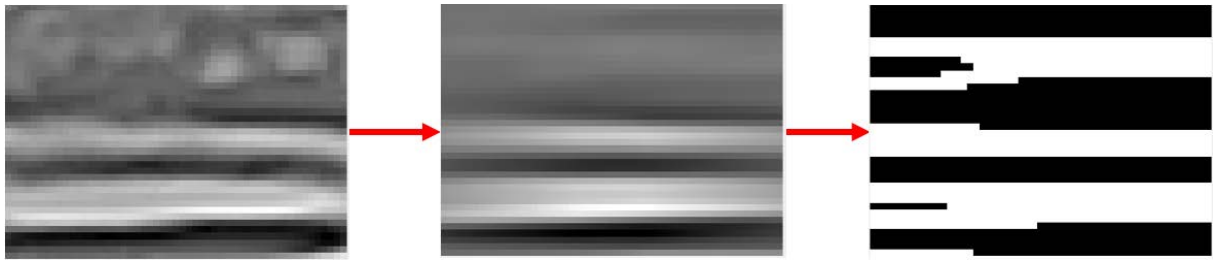


Figure 11: Motion filter application and binarization. To identify the intima, the sharpened image (left image) is processed with a motion filter, which mimics the linear motion of a camera (48). The filtered image (middle image) is subsequently binarized (right image) with Otsu's method.

Figure 12: Intima identification. The intima is identified in the binarized motion filtered image (left image) as the structure most closely above the adventitia. After identifying the intima (middle image) the upper border is determined (right image).

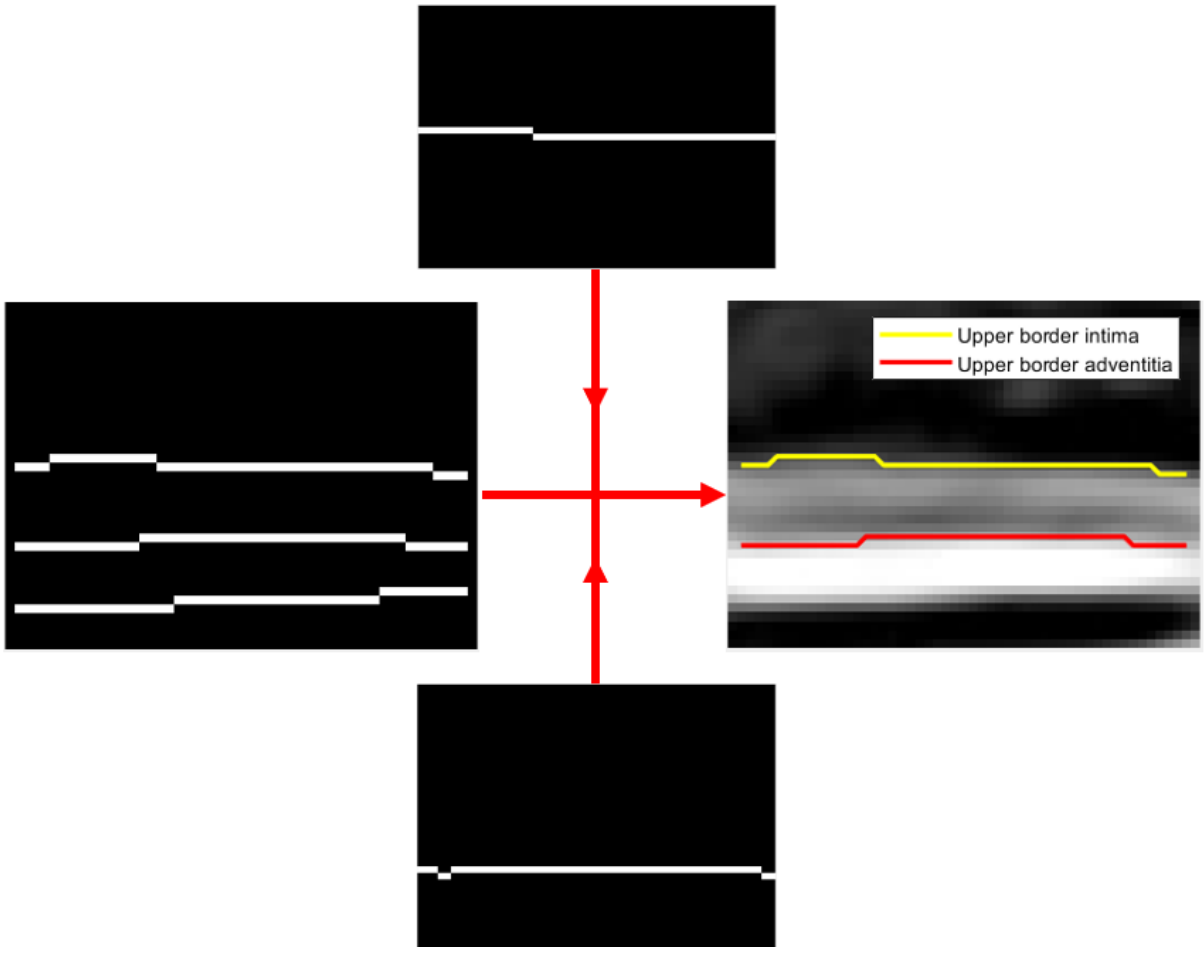


Figure 13: Upper borders adventitia and intima identification. The adventitial (bottom image) and intimal (top image) structures are used to identify the upper borders of the adventitia and intima in the binary edge image (left image). The upper borders are superimposed on the original ROI image (right image).

3.2 Pilot study

3.2.1 Research sample

The sample was predetermined to consist of ten healthy subjects, five men and five women. No subject was excluded based on being younger than 50 years or not providing a written informed consent. The five men and five women's mean ages were 62.2 and 55.6 years, respectively. The total number of US images that could have been acquired, is 240 (= 2 (sonographers) x 10 (subjects) x 2 (left/right side) x 3 (CTA / FTA / AX) x 2 (longitudinal / transversal US image)).

The number of US images actually acquired was 232. One female subject's images of the FTA were not acquired by the sonographers due to time constraints. Thus, the total number of US images acquired was 232 (= 240 – 2 (sonographers) x 1 (subject) x 2 (left/right side) x 1 (FTA) x 2 (longitudinal / transversal US image)).

3.2.2 Descriptive statistics of IMT measurements

The descriptive statistics of the manual longitudinal, semi-automatic longitudinal and transversally compressed IMT measurements of the CTA, FTA and AX, are shown in Table 8. The semi-automatic longitudinal measurements were highly reproducible, based on the very similar values between semi-automatic longitudinal measurement 1 and 2 for the CTA, FTA and especially the AX. When comparing the semi-automatic longitudinal measurements to the manual longitudinal measurements, the means of the semi-automatic longitudinal measurements were higher than those of the manual longitudinal measurements for the CTA and FTA. Conversely, the SDs of the semi-automatic longitudinal measurements were equal or smaller in comparison with the manual longitudinal measurements for the CTA, FTA and AX. The comparison between the transversally compressed and manual longitudinal measurements showed that the means for the CTA and FTA were similar, but for the AX, the means of the transversally compressed measurements were much higher than those of the manual longitudinal measurements. The SDs of the transversally compressed measurements were also higher than those of the manual longitudinal measurements, especially for the AX, with the only exception being the SD comparison for the CTA between sonographer B's transversally compressed and manual longitudinal measurements.

If the IMT could be measured in each image, then each measurement variable (i.e. semi-automatic longitudinal measurements 1 and 2, transversally compressed and manual longitudinal measurements A and B) would yield 40 measurements for each artery, except for the FTA, which would yield 38 measurements. Only the semi-automatic longitudinal and manual longitudinal measurements for the AX yielded 40 measurements, meaning that the IMT could always be determined in an AX longitudinal image. For every artery, sonographer B had a greater or equal amount of IMT measurements, compared to sonographer A. There were no images where sonographer A was able to measure the IMT and sonographer B was not. The semi-automatic algorithm was able to measure the IMT in an image, when sonographer B also was able to, except for one FTA image, where sonographer B could measure the IMT, but the semi-automatic algorithm failed (and sonographer A failed as well). When both sonographer A and B were not able to determine the IMT in an image, the semi-automatic algorithm also failed to do so.

Table 8: Descriptive statistics of the semi-automatic longitudinal, transversally compressed and manual longitudinal IMT measurements of the CTA, FTA, AX. The transversally compressed and manual longitudinal measurements were conducted by two sonographers, denoted as A and B. The semi-automatic measurements were conducted twice by a researcher. Explanation descriptive statistical parameters: min, smallest IMT measured; max, largest IMT measured; mean, mean IMT measured; SD, standard deviation of IMT measurements; n, number of IMT measurements.

IMT measurement (mm)	Min	Max	Mean	SD	N
CTA					
Semi-automatic longitudinal					
Measurement 1	0.21	0.41	0.30	0.06	38
Measurement 2	0.22	0.40	0.31	0.05	38
Sonographer A					
Transversally compressed	0.13	0.65	0.28	0.10	39
Longitudinal	0.17	0.41	0.28	0.06	37
Sonographer B					
Transversally compressed	0.18	0.42	0.28	0.06	39

Longitudinal	0.20	0.47	0.29	0.07	38
FTA					
Semi-automatic longitudinal					
Measurement 1	0.15	0.48	0.28	0.06	35
Measurement 2	0.15	0.47	0.29	0.06	35
Sonographer A					
Transversally compressed	0.13	0.48	0.25	0.07	35
Longitudinal	0.14	0.37	0.25	0.06	32
Sonographer B					
Transversally compressed	0.20	0.56	0.28	0.08	35
Longitudinal	0.10	0.47	0.27	0.07	36
AX					
Semi-automatic longitudinal					
Measurement 1	0.36	0.88	0.56	0.10	40
Measurement 2	0.36	0.88	0.56	0.10	40
Sonographer A					
Transversally compressed	0.21	1.18	0.72	0.20	34
Longitudinal	0.35	0.94	0.54	0.12	40
Sonographer B					
Transversally compressed	0.34	1.30	0.71	0.19	35
Longitudinal	0.41	0.99	0.57	0.12	40

3.2.3 Absolute agreement between IMT measurements

The absolute agreement between manual longitudinal, semi-automatic longitudinal and transversally compressed IMT measurements for the CTA, FTA and AX are shown in Table 9. The absolute agreement between semi-automatic and manual longitudinal measurements was moderate for the CTA and FTA, but good for the AX. The absolute agreement between transversally compressed and manual longitudinal measurements was poor for the CTA, FTA and AX, and all their respective 95% CIs included 0.00, indicating that the absolute agreement was not significant. The absolute agreement of the manual longitudinal measurements between sonographer A and B was moderate for the CTA and FTA, but good for the AX. The absolute agreement of the transversally compressed measurements between sonographer A and B was moderate for the CTA, FTA and AX. Lastly, the absolute agreement between the semi-automatic longitudinal measurements was very good for the CTA, FTA and AX.

Table 9: Absolute agreement between manual longitudinal, semi-automatic longitudinal and transversally compressed IMT measurements for the CTA, FTA and AX. The absolute agreement was operationalized with the ICC and its 95% CI. The manual longitudinal and transversally compressed measurements were conducted by two sonographers, denoted as A and B. The semi-automatic measurements were conducted twice by a researcher.

	CTA	n	FTA	n	AX	n
Manual longitudinal A versus						
Semi-automatic longitudinal 1 (95% CI)	0.75 (0.41-0.88)	37	0.56 (0.18-0.78)	32	0.81 (0.67-0.90)	40
Semi-automatic longitudinal 2 (95% CI)	0.73 (0.35-0.88)	37	0.56 (0.16-0.78)	32	0.81 (0.67-0.89)	40
Transversally compressed (95% CI)	0.27 (0.00-0.54)	37	0.25 (0.00-0.54)	32	0.14 (0.00-0.40)	34
Manual longitudinal B versus						
Semi-automatic longitudinal 1 (95% CI)	0.66 (0.43-0.81)	38	0.70 (0.48-0.83)	35	0.89 (0.81-0.94)	40
Semi-automatic longitudinal 2 (95% CI)	0.66 (0.44-0.81)	38	0.67 (0.44-0.82)	35	0.89 (0.80-0.94)	40
Transversally compressed (95% CI)	0.22 (0.00-0.51)	38	0.07 (0.00-0.39)	35	0.09 (0.00-0.36)	35
Sonographer A versus B						

Longitudinal (95% CI)	0.57 (0.30-0.75)	37	0.60 (0.32-0.78)	32	0.87 (0.73-0.93)	40
Transversally compressed (95% CI)	0.52 (0.25-0.72)	39	0.63 (0.34-0.80)	35	0.71 (0.50-0.85)	34
Semi-automatic longitudinal						
Measurement 1 versus 2 (95% CI)	1.00 (0.99-1.00)	38	0.99 (0.98-1.00)	35	1.00 (1.00-1.00)	40

3.2.4 Limits of agreement between IMT measurements

The Bland-Altman plots of the differences between the semi-automatic longitudinal and manual longitudinal IMT measurements are shown for the CTA in Figure 14. Semi-automatic measurements 1 and 2 were compared to sonographer A (top plot) and sonographer B's (bottom plot) manual measurements. In both plots, the orange dots almost completely cover each blue dot and the measurement bias lines of the first and second semi-automatic measurements are also very close to each other. The semi-automatic measurements were therefore highly reproducible. The bias lines show a slight overestimation bias of the semi-automatic measurements relative to the manual longitudinal measurements in both plots, but they are well below the upper statistical limit of agreement. One first and second semi-automatic measurement clearly deviated beyond the upper statistical limit of agreement in both plots, but no semi-automatic measurement deviated beyond the clinical limits of agreement.

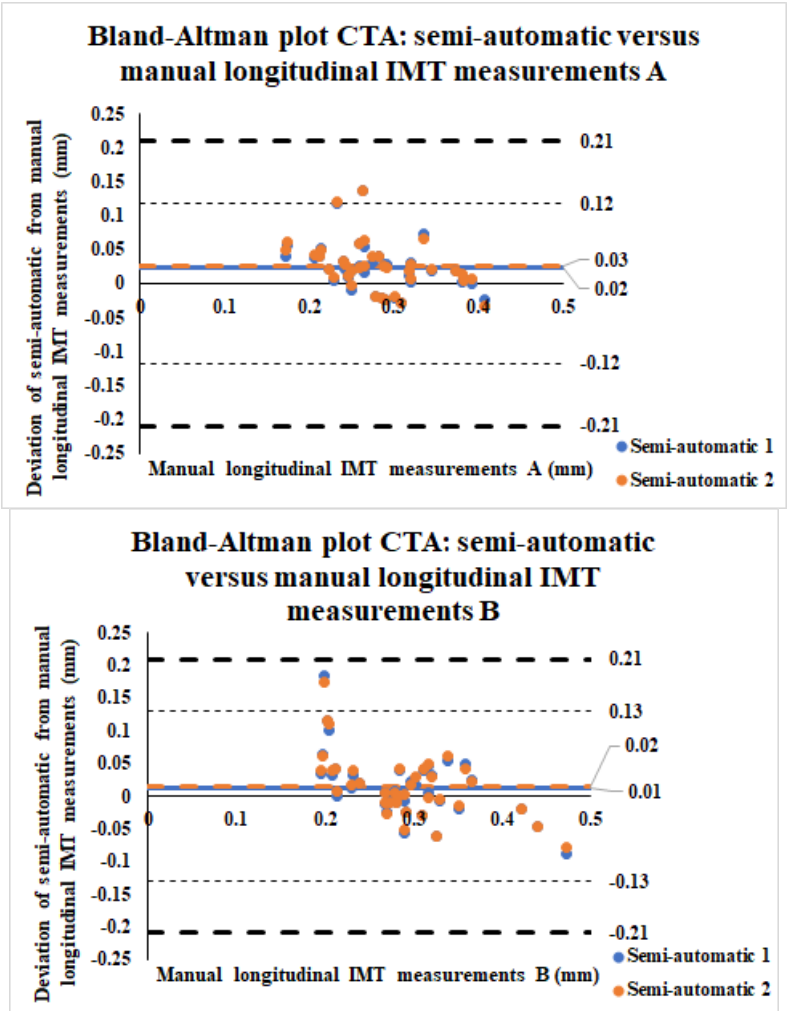


Figure 14: Bland-Altman plots of the differences between the semi-automatic longitudinal and manual longitudinal IMT measurements for the CTA. The first and second semi-automatic measurements were compared to sonographer A (top plot) and sonographer B's (bottom plot) manual measurements. The horizontal axis represents the manual measurements and the vertical axis the deviation of the semi-

automatic from the manual measurements (manual measurements – semi-automatic measurements). The blue line and orange dashed line represent the measurement bias of semi-automatic measurement 1 and 2, respectively. The bias was operationalized with the mean deviation of the semi-automatic measurements from the manual measurements. The thinner and thicker dashed black lines represent the statistical and clinical limits of agreement, respectively. The statistical limits of agreement were set to +/- 2 SDs of the manual measurements. The clinical limits of agreement were based on the clinical IMT thresholds in Table 1, and set to half of the interval between a 'normal' and a 'pathological' IMT. The clinical limits of agreement for the IMT measurements in the CTA, FTA and AX were respectively: +/- 0.21 mm, +/- 0.175 mm and +/- 0.45 mm.

Analogous plots to Figure 14 are shown for the FTA in Figure 15. The semi-automatic measurements again showed a high reproducibility. The bias lines of the semi-automatic measurements again have a slight overestimation bias, but they are well within the upper statistical limit of agreement. The upper and lower statistical limits of agreement were each crossed by a first and second semi-automatic measurement in the top plot compared with sonographer A's measurements, but all semi-automatic measurement deviations remained within the clinical limits of agreement. When the semi-automatic measurements are compared to sonographer B's measurements in the bottom plot, it can be seen that all semi-automatic measurements deviated less than the statistical limits of agreement, except for one first and second measurement, which deviated beyond the upper statistical and clinical limit of agreement. Sonographer A's measurement in the same US image deviated only 0.09 mm from sonographer B's measurement, below the upper clinical and statistical limit of agreement. This indicates that this particular deviation of the semi-automatic measurement was also outside the range of the manual measurements of sonographer A and B. This means that if sonographer B had assessed an FTA IMT to be non-pathological, the semi-automatic algorithm could have found it to be pathological. The analogous plots for the AX are shown in Figure 16. The semi-automatic measurements again show high reproducibility. The semi-automatic measurements have a slight overestimation bias in relation to sonographer A's measurements and a slight underestimation bias in relation to sonographer B. No semi-automatic measurement deviated beyond the statistical and clinical limits of agreement.

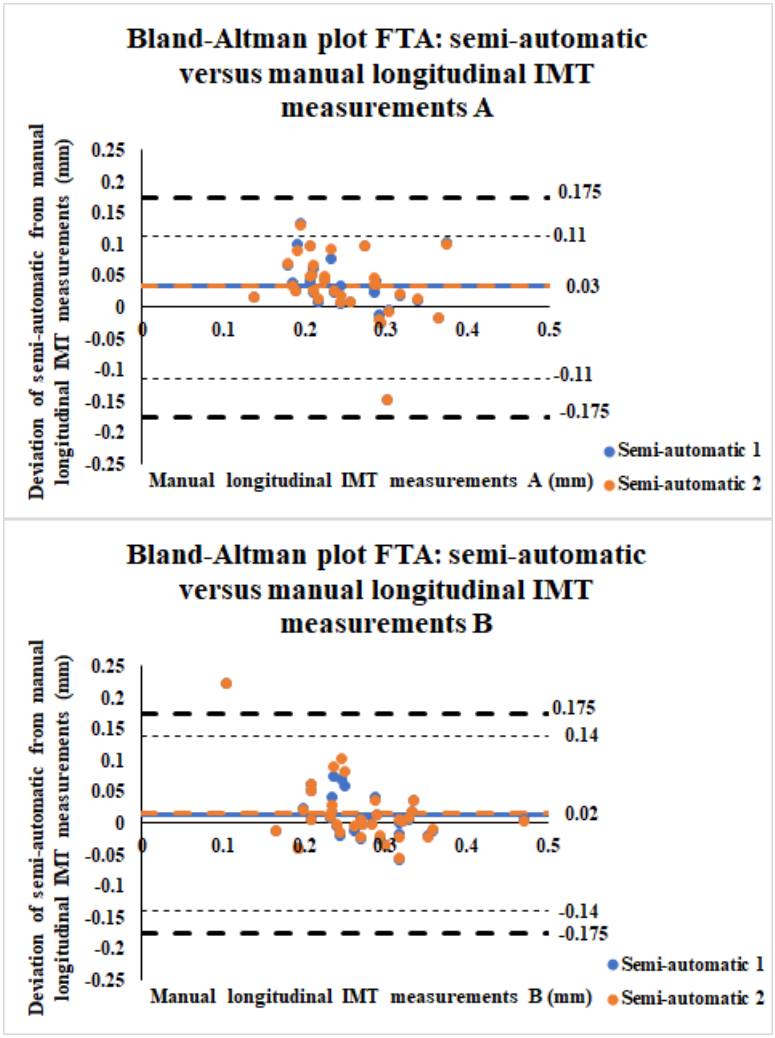


Figure 15: Bland-Altman plots of the differences between the semi-automatic longitudinal and manual longitudinal IMT measurements for the FTA. The first and second semi-automatic measurements were compared to sonographer A (top plot) and sonographer B's (bottom plot) manual measurements. See the caption of Figure 14 for the definitions of the axes and horizontal lines.

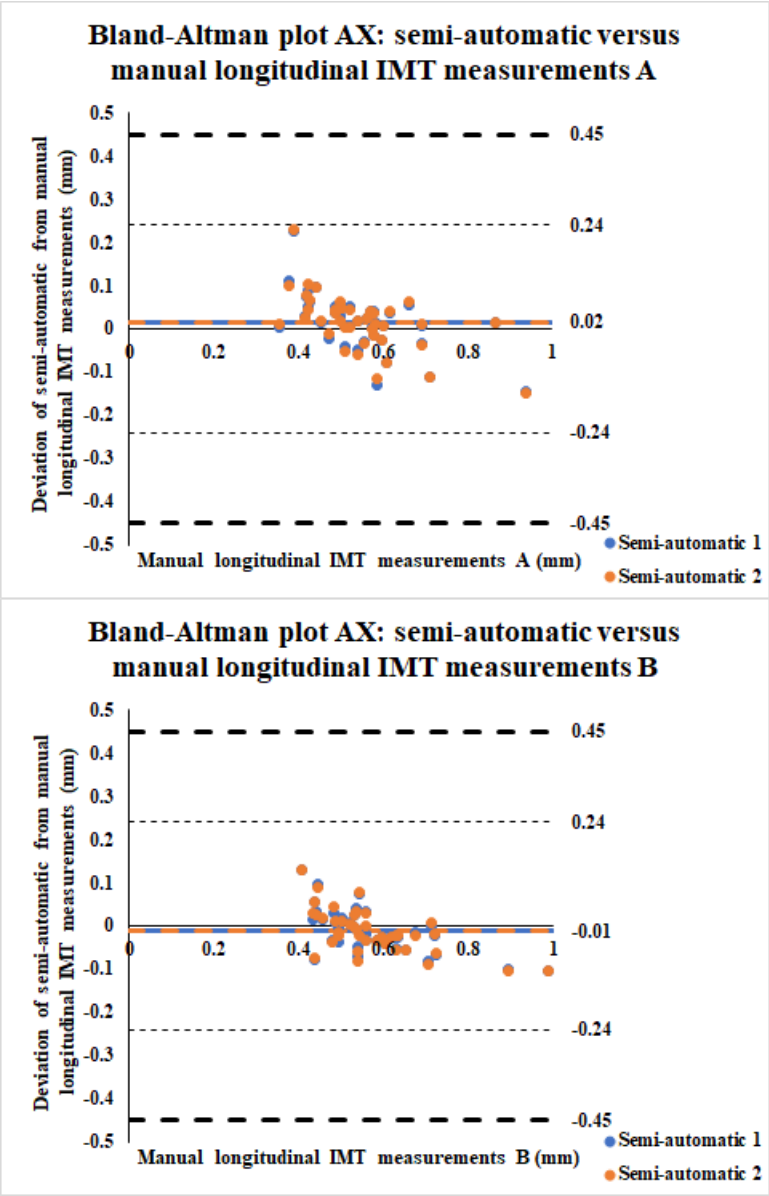
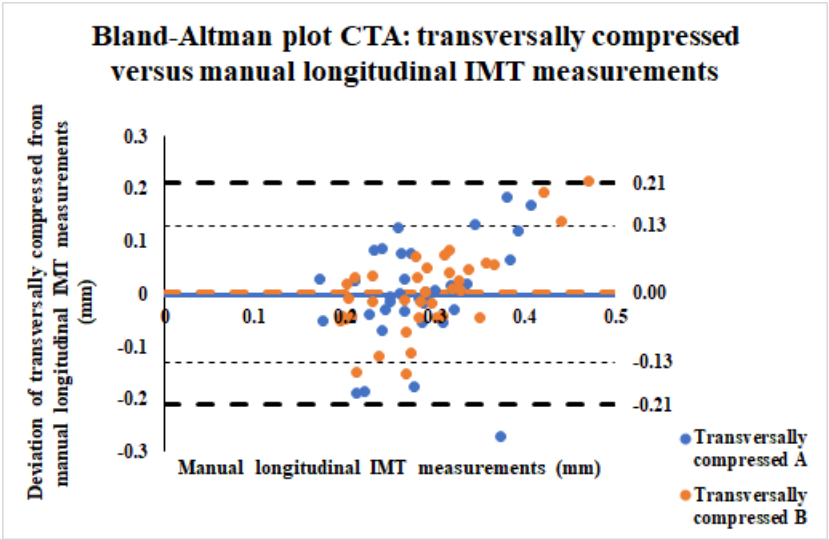


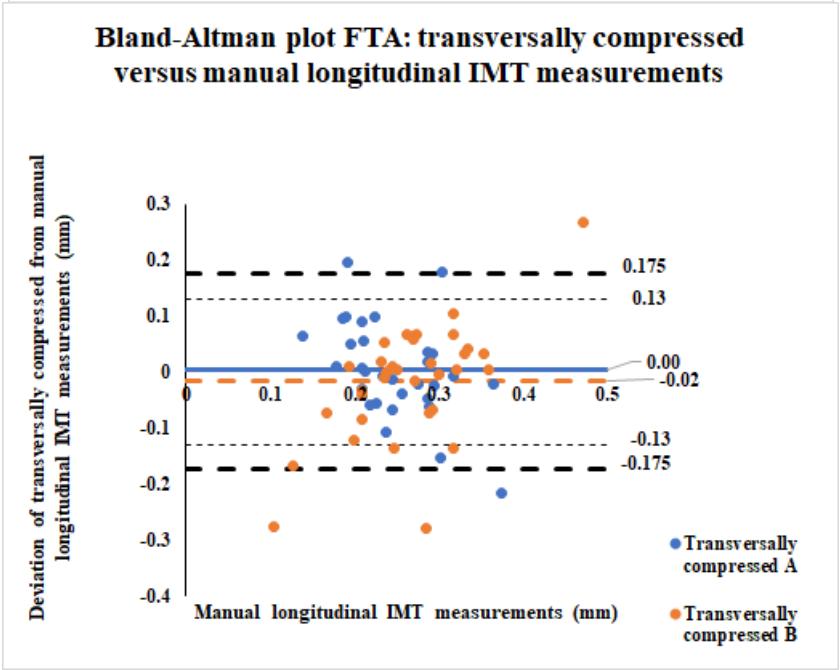
Figure 16: Bland-Altman plots of the differences between the semi-automatic longitudinal and manual longitudinal IMT measurements for the AX. The first and second semi-automatic measurements were compared to sonographer A (top plot) and sonographer B's (bottom plot) manual measurements. See the caption of Figure 14 for the definitions of the axes and horizontal lines.

The Bland-Altman plots of the differences between the transversally compressed and manual longitudinal IMT measurements for the CTA, FTA and AX are shown in Figure 17. In the CTA plot, sonographer A and B's measurement bias lines are both 0.00. This means the transversally compressed measurements show neither an underestimation nor overestimation bias with respect to the manual longitudinal measurements. There are several transversally compressed measurements that deviate beyond the statistical limits of agreement, but only one measurement clearly deviates beyond the clinical limits of agreement. This was sonographer A's measurement that crossed the lower clinical limit of agreement. This means that if sonographer A had assessed a CTA IMT to be pathological in a longitudinal US image, he could have found it to be non-pathological in the corresponding transversal image with vessel compression.

The FTA plot shows that sonographer A's transversally compressed measurements neither has an underestimation nor overestimation bias, while sonographer B's transversally compressed

measurements have a slight underestimation bias with respect to the manual longitudinal measurements. Two measurements (one of sonographer A and one of B) also deviate beyond the upper clinical limit of agreement and three measurements (one of sonographer A and two of B) beyond the lower clinical limit of agreement. Lastly, the AX plot shows two bias lines well below the horizontal axis, which implies that the transversally compressed measurements have a significant underestimation bias. This is illustrated by the great majority of deviations that are clustered below the horizontal axis. Still, four transversally compressed measurements (two of sonographer A and two of B) deviated beyond the upper statistical limit of agreement, but remained below the upper clinical limit of agreement. Conversely, 18 measurements (12 of sonographer A and six of B) deviated beyond the lower statistical limit of agreement. Five measurements (three of sonographer A and two of B) deviated beyond the lower clinical limit of agreement.





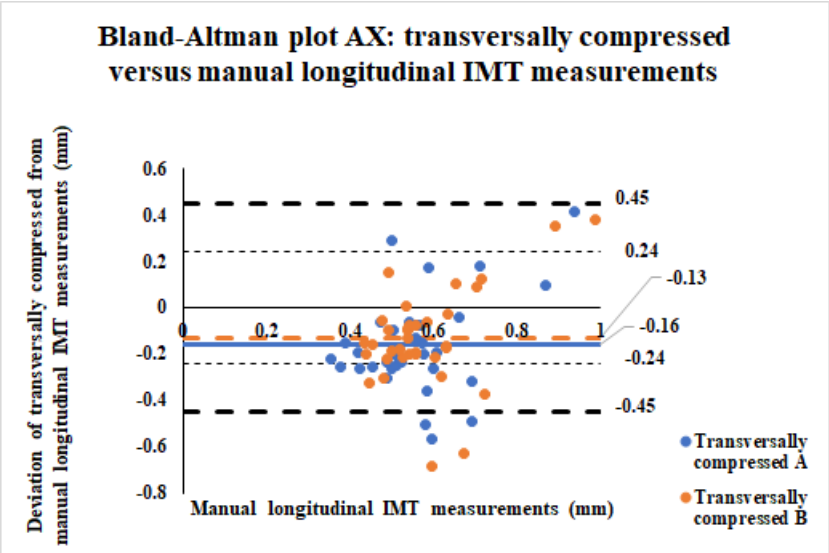


Figure 17: Bland-Altman plots of the differences between the transversally compressed and manual longitudinal IMT measurements for the CTA (top plot), FTA (middle plot) and the AX (bottom plot). See the caption of Figure 14 for the definitions of the axes and horizontal lines.

4 Discussion

4.1 Retrospective cohort study

The sensitivity of US in diagnosing GCA was lower than TAB, but still comparable to it (75.8% versus 80.8%). US and TAB had an NPV of 89.3% and 80.0%, respectively, and both had a specificity and PPV of 100%, meaning that a positive result of either examination would always lead to a GCA diagnosis. Performing TAB might therefore be unnecessary for a GCA diagnosis when the US result is positive. This result aligns with the recommendations of the European League Against Rheumatism (EULAR) for the use of imaging large vessel vasculitis in clinical practice. EULAR state that US is the first choice for an early imaging test of patients with a high pre-test probability of GCA, with a positive US result being sufficient for a GCA diagnosis. (12) One of the reasons why US is the first choice is its low-cost, both financially and clinically, as was illustrated by the fact that US was performed far more frequently in the sample than TAB (100 US versus 46 TAB examinations). The concept of high pre-test probability is also pertinent in the explanation of the NPV difference between US and TAB. Predictive values primarily depend on the pre-test probability of having the disease (49). Patients suspected of GCA referred to the Rheumatology Department already have a higher pre-test probability of GCA than the general population, but only those with the highest pre-test probability undergo TAB. The pre-test probability could be so high, that even a negative TAB result would not be enough to remove this suspicion. The lower NPV of TAB could therefore be a reflection of the difference in pre-test probability between patients who undergo TAB and US, and again accentuate the low-cost of using US.

The sensitivity and specificity of US were higher than those found by Luqmani et al. (54% and 81%, respectively) (2), which may be explained by the sonographers' greater experience in the current study. Another possible explanation for this discrepancy is the methodological difference in defining the reference diagnosis. The reference diagnosis of the current study was based on all the available information (including the US results) used by the physician to either diagnose or exclude GCA. Luqmani et al. purposefully prevented the US results from being used to establish the reference diagnosis. If they had not done this, then the information gained from the US results could have caused the reference diagnosis to be more aligned with the US results, thereby (artificially) increasing the sensitivity and specificity of US. This issue

highlights the main limitation of the current study: the absence of an independent gold standard for diagnosing GCA. Every independent variable analyzed in the current study was taken into account when determining the reference diagnosis. Thus, analyzing the relationship between each independent variable and the reference diagnosis is inherently circular to a degree. This is most clearly demonstrated with the results of US and TAB. The fact that each positive result led to a GCA diagnosis was to be expected, as no physician would ignore a positive result when diagnosing GCA. This circular reasoning diminishes the validity of the results, but is inevitable in the absence of an independent gold standard for diagnosing GCA.

A positive halo sign, compression sign and a 'pathological' IMT were, as expected, very significantly associated with GCA patients, but they were also present in 7.5%, 3.0% and 3.0% of non-GCA patients, respectively. These three US characteristics are considered highly suggestive of GCA, but sometimes their presence is dubious or false-positive due to other causes, such as the natural thickening of an arterial wall in older patients (2). In total there were 25 positive US results and 30 positive halo signs, which shows that the sonographer judged five positive halo signs not to be related to GCA. This speaks to the importance of the sonographer's expertise in assessing US characteristics of GCA. A 'normal' IMT was significantly related to a negative US result, but a 'likely pathological' IMT did not show a significant relationship with a positive US result. In fact, even though not significant, a higher percentage of patients with a negative US result had a 'likely pathological' IMT than those with a positive US result. 'Likely pathological' is therefore a misnomer, with 'possibly pathological' being a better name for the threshold. There are no universally agreed upon thresholds for the clinical assessment of the IMT, with this result further corroborating that fact. The clinical assessment of the IMT has diagnostic value, if it is either 'normal' or 'pathological', but any IMT in between is more difficult to interpret, necessitating the physician to rely on other clinical findings. The halo sign, compression sign and IMT can be assessed for each TA branch and AX individually, which were aggregated into an overall halo sign, compression sign and IMT in the current study. An analysis of the US characteristics of each individual artery could not be conducted due to missing values and non-systematic reporting of US examination findings. Often only the presence of US characteristics was noted without entailing the artery/branch or laterality. These limitations impeded further investigation into the number and location of US characteristics and their relationship with the diagnosis of GCA, which was shown to be important by other studies (31,33).

The sensitivity of TAB was considerably higher in the current study than the 39% found in the study of Luqmani et al. This discrepancy could be explained by the fact that novice surgeons conducted the TAB in their study. This increased the risk of TAB yielding the wrong tissue structures or wrong TA segments, as GCA is known to have 'skip lesions' that often do not affect an entire artery. Each patient in their sample (n=381) underwent TAB, of which 13% did not yield an artery. Moreover, 43% of the TAB examinations yielded an artery shorter than 1 cm, which is the minimum TAB length recommended by the British Society for Rheumatology (2). The risk of these errors was lower for the current study, because the sonographer marked the skin area of the pathological TA segment before TAB. Of the 46 TAB examinations performed in the current study, only one TAB did not yield an artery (but a vein), and no TAB lengths were less than 1 cm. Another issue raised by Luqmani et al. in explaining the low TAB sensitivity was the potential masking effect of corticosteroid use. The results of the current study did not show the anti-inflammatory effects of corticosteroids to significantly mask the presence of GCA during US and TAB examinations. Relatively more patients with negative US results used corticosteroids than those with positive results, which aligns with a masking effect, but the difference was not significant. The p-value of the TAB difference for corticosteroid use

was almost significant (0.062), but the corticosteroid use was actually relatively more frequent for positive TAB results than negative ones. A possible explanation is that the pre-test probability of GCA was already so high that corticosteroids were prescribed before TAB was conducted. The main limitation of this analysis was that the corticosteroid dose and length of use were not taken into account, which was again due to missing values and non-systematic reporting.

The following demographic and clinical characteristics were found to be significantly more associated with GCA than non-GCA patients: female gender, jaw claudication, TA pain, thickening and an impalpable pulse, and elevated ESR and CRP values. Three GCA patients (8.8% of all GCA patients) in the current study had both a negative US result and a negative TAB result. However, they did have most, if not all, of the aforementioned significant demographic and clinical characteristics. It remains difficult to diagnose GCA without the use of imaging and/or histology, but the totality of demographic and clinical characteristics is important in determining an early pre-test probability of GCA. Laskou et al. (2019) developed a GCA probability score to aid the early differentiation of newly suspected GCA patients into a high-probability and low-probability group. The sensitivity and specificity of the GCA probability score were 95.7% and 86.7%, respectively, compared to the documented GCA diagnosis after six-months follow-up. Combined with US, Laskou et al. propose the following algorithm for early GCA diagnosis: a low-probability score and negative US result exclude GCA; a high-probability score and positive US result confirm GCA; conflicting probability scores and US results require TAB and/or further imaging. (50) The pre-test probability score of Laskou et al. has yet to be externally validated, but it does underline the importance of integrating US in the overall strategy of diagnosing GCA.

4.2 Pilot study

The semi-automatic IMT measurements had a good absolute agreement with the manual longitudinal measurements for the AX, but a moderate absolute agreement for the CTA and FTA. These absolute agreement values are similar to those between sonographer A and B's manual longitudinal measurements, which implies that the semi-automatic IMT measurements are as valid as manual measurements. The absolute agreement between the transversally compressed measurements and the manual longitudinal measurements was poor and statistically insignificant for the CTA, FTA and AX. This means that the transversal compression technique is not a valid method for IMT measurement. The absolute agreement between the first and second semi-automatic measurements was almost perfect for the CTA, FTA and AX, indicating that the semi-automatic measurements are highly reliable. The absolute agreement between sonographer A and B's transversally compressed measurements was similar to the absolute agreement between the manual longitudinal measurements for the CTA and FTA, but lower for the AX. Thus, the transversal compression technique is as reliable as manual longitudinal measurements for the CTA and FTA.

The greater absolute agreement for the AX between the semi-automatic and manual longitudinal measurements was expected, because the AX is a much larger artery than the CTA and FTA. A measurement discrepancy in a large artery has a smaller impact on the absolute agreement than the same discrepancy in a small artery. This could also explain why the reliability of the transversally compressed measurements was higher for the AX, then for the CTA and FTA. The stark difference in absolute agreement values with the manual longitudinal measurements between the semi-automatic and transversally compressed measurements can be explained by the fact that both manual longitudinal and semi-automatic measurements were performed in the same longitudinal images, while the transversally compressed measurements

were performed in transversal images. Although care was taken to measure the IMT of the same arterial segment in both longitudinal and transversal images, it was virtually impossible to accomplish in practice due to the limited dexterity allowed by the probe and the slippery nature of US gel. Moreover, accurately measuring the IMT in transversal images with vessel compression was difficult after the examination due to the ambiguous appearance of the compressed artery. Even though marks were placed in each transversal image to denote the ROI for the IMT measurement, the orientation in which the artery had compressed remained ambiguous. Pressure application with the probe does not always lead to vessel compression in the same direction, as it can also make the artery move beneath the probe, which may change the angle of pressure application. Additionally, nearby structures can be difficult to distinguish from the compressed intima-media, adding to the ambiguity. When a sonographer compresses the artery during US examination, he can clearly see how the artery deforms and how the IMT should be measured. The sonographers in the current study measured the IMT in anonymized transversal images at least several days after the examination, and therefore did not have this information. Recording videos of the transversal compressions could have mitigated this issue. Lastly, the transversal compression technique assumes that the arterial near and far walls are symmetrical, which does not always hold true, especially not for GCA patients with eccentric (i.e. asymmetrical) arterial wall thickening. The mean IMT between the near and far wall would therefore invariably differ from the IMT of only the far wall.

The results of the Bland-Altman plots show that the semi-automatic longitudinal and manual longitudinal IMT measurements have a good clinical agreement for the CTA, FTA and AX. The Bland-Altman plots of the differences between the transversally compressed and manual longitudinal IMT measurements show that the deviations were generally greater and more dispersed than those between the semi-automatic longitudinal and manual longitudinal IMT measurements. However, the transversally compressed measurement for the AX did show a significant underestimation bias. AXs are inherently difficult to compress due to their size. Applying just enough pressure to make the Doppler signal disappear was therefore difficult to achieve. As a result, the sonographers often overcompressed the AX, which could have led to an underestimation bias of the IMT. The number of transversally compressed measurements that deviated beyond the clinical limits of agreement was one, five and five for the CTA, FTA and AX, respectively. These transversally compressed measurements and their corresponding transversal images did not differ ostensibly from the other measurements and images with deviations within the clinical limits of agreement. It could be that the aforementioned validity and reliability issues concerning the transversal compression technique were compounded in these specific measurements and that actually all transversally compressed measurements suffered from these issues to a certain degree.

No semi-automatic measurement deviated beyond any clinical limit of agreement in the Bland-Altman plots, except for one first and second measurement for the FTA, which deviated beyond the upper clinical limit of agreement when compared to sonographer B's measurements. Sonographer A's measurement in the same image was also larger than sonographer B's measurements, but did not deviate beyond the upper clinical limit of agreement, indicating that the deviation of the semi-automatic measurements was outside the range of manual measurements. However, inspecting each measurement in this specific FTA image, it becomes clear that the agreement between sonographer A and B is illusory, because both measured completely different structures in the image. The semi-automatic algorithm measured the same structure as sonographer A, but included a larger part of it. This means that these semi-automatic measurements were not less valid than the manual measurements, because there was no agreement between the sonographers which structure constituted the intima-media. The

difference in image interpretation between sonographer A and B could also be seen in their discrepancy in the number of IMT measurements. Sonographer B had a greater or equal amount of IMT measurements for every artery, compared to sonographer A. Sonographer B was always able to measure the IMT in an image, when sonographer A was also able to do so. There is no objective criterion to judge if the IMT is actually measurable in an image. It is important to note that neither sonographer was sure if his assessment of these contested images was correct. This underlines how difficult image interpretation can be, even for experienced sonographers, when the layers of the arterial wall are poorly visible.

Several automatic operations of the semi-automatic algorithm, such as Canny edge detection and Otsu's method of binarization, take into account all intensity values in the ROI to calculate their thresholds. This means that variation in the ROI window can influence the IMT measurements, if it significantly changes the overall intensity profile of the ROI. However, the high reliability results show that the semi-automatic algorithm is robust against such effects. The main limitation of the semi-automatic algorithm is the need for it to be guided by the manual ROI selection. In the current study each IMT measurement location was marked in the image, which greatly reduced the variability in the ROI selection, but was necessary to establish the validity and reliability of the semi-automatic algorithm. In the end, the semi-automatic algorithm is a tool to assist the sonographer in measuring the IMT, and cannot serve as a substitute for the sonographer's expertise, which are not only necessary to select the ROI, but also to evaluate the result. Nevertheless, even though there is an inherent variability in image interpretation, the results of the current study have shown that it can be partly reduced by performing IMT measurements semi-automatically. Moreover, the semi-automatic algorithm could be expanded to provide more information on the arterial wall, such as the largest and smallest IMT measured, and the relative intensity values of each layer, which may not only provide additional insights for sonographers to assess and diagnose GCA, but may also be applicable to arterial diseases, such as atherosclerosis.

Important limitations of the current study were the research sample and the use of manual longitudinal measurements as the benchmark. The research sample only consisted of healthy subjects, who did not have enlarged IMTs and were therefore less representative of GCA patients. Both sonographers noted how challenging it was to image a measurable IMT segment of the CTA and FTA longitudinally. The intima-media in the transversal images with compression was also more ambiguous than in GCA patients. These difficulties stem from the fact that normal TA intima-media layers are thin and they blend easily into the image background. Analogous to the aforementioned differences between larger and smaller arteries, measurement differences in larger pathological IMTs have a smaller impact on the absolute agreement than differences in smaller normal IMTs. The absolute agreement values of the semi-automatic, transversally compressed and manual longitudinal measurements could have therefore been higher, if the research sample would have consisted of GCA patients. Lastly, both semi-automatic and transversally compressed measurements were compared with manual longitudinal measurements. This does not mean that manual longitudinal measurements are the gold standard, because directly measuring the IMT in vivo is currently not possible. If future research would solve the other limitations of the current study and find that the transversally compressed measurements were reliable, but still poorly agreeable with the manual longitudinal measurements, then one could question if the longitudinal measurements are the most accurate approximation of the real IMT.

In the end, what is paramount, is the clinical significance of each IMT measurement method. To investigate this matter, one could conduct a study with two sonographers and a sample of

patients suspected of GCA, whose medical history are hidden from the sonographers. Each sonographer would examine every patient independently, measuring the IMT manually longitudinally, semi-automatically longitudinally and transversally compressed, for every artery, only seeing the measurement visually without being shown their quantitative values. Every examination would generate three US results, each one based on a single IMT measurement method for every artery. Ideally, each patient would be diagnosed without using US. These three US results per examination would then be compared with the diagnosis to determine the US diagnostic value per measurement method. If either the semi-automatic algorithm or transversal compression technique truly adds value to the diagnosis of GCA, then its US result should have a higher sensitivity and/or specificity than the US result of the manual longitudinal IMT measurements. Each sonographer would also repeat the measurements in the same images (accompanied by supporting videos of transversal images with compression) a week later (without knowing the diagnosis), which would allow one to see if the intrarater reliability of either the semi-automatic or transversally compressed measurements is higher than the intrarater reliability of longitudinal manual measurements and if this difference would be clinically significant. Lastly, the two sonographers' diagnostic and measurement results are compared to see if the diagnostic results are robust and if the differences in the interrater reliabilities of the IMT measurement methods are clinically significant.

5 Conclusion

US of the TA and AX can be of added value in diagnosing GCA early in suspected patients, with a sensitivity (75.8%) and specificity (100%) comparable to TAB (80.0 and 100%, respectively). A positive halo sign, positive compression sign and 'pathological' IMT are all strongly associated with a positive GCA diagnosis, but they were also present in 7.5%, 3.0% and 3.0% of non-GCA patients, respectively. The clinical assessment of the IMT has diagnostic value, if it is either 'normal' or 'pathological', but any IMT in between is more difficult to interpret, necessitating the physician to rely on other clinical findings. The relationship between each independent variable and the reference diagnosis is inherently circular to a degree, which diminishes the validity of the results, but is inevitable in the absence of an independent gold standard for diagnosing GCA. 8.8% of all GCA patients had both a negative US result and a negative TAB result, but were diagnosed based on demographic and clinical characteristics. It remains difficult to diagnose GCA without the use of imaging and/or histology, but the totality of demographic and clinical characteristics is important in determining an early pre-test probability of GCA. This underlines the importance of integrating US in the overall strategy of diagnosing GCA.

The semi-automatic algorithm developed to measure IMT in longitudinal US images is primarily based on Canny edge detection, with the manual component pertaining to the selection of the ROI in images. The semi-automatic IMT measurements were found to have a moderate absolute agreement with the manual longitudinal measurements for the CTA and FTA, but a good absolute agreement for the AX. These absolute agreement values were similar to those between sonographer A and B's manual longitudinal measurements. The clinical agreement between semi-automatic and manual longitudinal measurements was good for every artery. Thus, semi-automatic measurements are as valid as manual longitudinal measurements and highly reliable for the CTA, FTA and AX. The absolute agreement between the transversally compressed measurements and the manual longitudinal measurements was poor and statistically insignificant for the CTA, FTA and AX. This means that the transversal compression technique is not a valid method for IMT measurement. The reliability of transversally compressed measurements is comparable to manual longitudinal measurements for the CTA and FTA, but lower for the AX. Future research has to be conducted with a sample

of patients suspected of GCA, and relate the IMT measurements to the GCA diagnosis in order to fully explore the potential of the semi-automatic algorithm and transversal compression technique to increase the value of US in diagnosing GCA early, which is necessary to minimize both the GCA and corticosteroid complications of patients suspected of GCA.

6 References

1. Hakvoort L, Dubbeld P, Ballieux M, Dijkstra R, Meijman H, Weisscher P, et al. NHG-Standaard Polymyalgia rheumatica en arteriitis temporalis. Huisarts Wet. 2010;53(2):88–98.

2. Luqmani R, Lee E, Singh S, Gillett M, Schmidt WA, Bradburn M, et al. The role of ultrasound compared to biopsy of temporal arteries in the diagnosis and treatment of giant cell arteritis (TABUL): A diagnostic accuracy and cost-effectiveness study. Health Technol Assess (Rockv). 2016;

3. Yates M, MacGregor AJ, Robson J, Craven A, Merkel PA, Luqmani RA, et al. The association of vascular risk factors with visual loss in giant cell arteritis. Rheumatology. 2016;(December 2016):kew397.

4. Petri H, Nevitt A, Sarsour K, Napalkov P, Collinson N. Incidence of giant cell arteritis and characteristics of patients: Data-driven analysis of comorbidities. Arthritis Care Res. 2015;

5. Ninan J, Lester S, Hill C. Giant cell arteritis. Best Pract Res Clin Rheumatol. 2016;

6. Brouwer E, Boumans D, Slart R, Rutgers B, Mulder U, Colin E, et al. Reuscelarteriitis; systeemvasculitis van de grote vaten. FocusVasculair. 2018;September(3).

7. Borchers AT, Gershwin ME. Giant cell arteritis: A review of classification, pathophysiology, geoepidemiology and treatment. Autoimmunity Reviews. 2012.

8. Ciccia F, Rizzo A, Ferrante A, Guggino G, Croci S, Cavazza A, et al. New insights into the pathogenesis of giant cell arteritis. Autoimmunity Reviews. 2017.

9. Samson M, Corbera-Bellalta M, Audia S, Planas-Rigol E, Martin L, Cid MC, et al. Recent advances in our understanding of giant cell arteritis pathogenesis. Autoimmunity Reviews. 2017.

10. Falardeau J. Giant Cell Arteritis. Neurologic Clinics. 2010.

11. Monti S, Floris A, Ponte C, Schmidt WA, Diamantopoulos AP, Pereira C, et al. The use of ultrasound to assess giant cell arteritis: review of the current evidence and practical guide for the rheumatologist. Rheumatology. 2017;

12. Dejaco C, Ramiro S, Duftner C, Besson FL, Bley TA, Blockmans D, et al. EULAR recommendations for the use of imaging in large vessel vasculitis in clinical practice. Ann Rheum Dis. 2018;

13. Kermani TA, Schmidt J, Crowson CS, Ytterberg SR, Hunder GG, Matteson EL, et al. Utility of Erythrocyte Sedimentation Rate and C-Reactive Protein for the Diagnosis of Giant Cell Arteritis. Semin Arthritis Rheum. 2012;

14. Ninan J V., Lester S, Hill CL. Giant cell arteritis: beyond temporal artery biopsy and steroids. Internal Medicine Journal. 2017.

15. Berger CT, Sommer G, Aschwanden M, Staub D, Rottenburger C, Daikeler T. The clinical benefit of imaging in the diagnosis and treatment of giant cell arteritis. Swiss Med Wkly. 2018;

16. Allison MC, Gallagher PJ. Temporal artery biopsy and corticosteroid treatment. Ann Rheum Dis. 1984;

17. Meiburger KM, Acharya UR, Molinari F. Automated localization and segmentation techniques for B-mode ultrasound images: A review. Computers in Biology and Medicine. 2018.

18. Schmidt WA, Blockmans D. Investigations in systemic vasculitis – The role of imaging. Best Practice and Research: Clinical Rheumatology. 2018.

19. Powles AE, Martin DJ, Wells IT, Goodwin CR. Physics of ultrasound. Anaesthesia and Intensive Care Medicine. 2018.

20. Aldrich JE. Basic physics of ultrasound imaging. Critical Care Medicine. 2007.

21. Drost WT. Physics of Ultrasound Imaging. In: Textbook of Veterinary Diagnostic Radiology. 2018.

22. Schmidt WA, Seifert A, Gromnica-ihle E, Krause A, Natusch A. Ultrasound of proximal upper extremity arteries to increase the diagnostic yield in large-vessel giant cell arteritis. Rheumatology. 2008;

23. Schmidt WA. Ultrasound in the diagnosis and management of giant cell arteritis. Rheumatol (United Kingdom). 2018;

24. Weyand CM, Liao YJ, Goronzy JJ. The immunopathology of giant cell arteritis: Diagnostic and therapeutic implications. Journal of Neuro-Ophthalmology. 2012.

25. Koster MJ, Matteson EL, Warrington KJ. Giant Cell Arteritis. Hear Rheum Autoimmune Inflamm Dis Pathophysiol Clin Asp Ther Approaches [Internet]. 2017;63(3):367–87. Available from: http://dx.doi.org/10.1016/j.disamonth.2016.10.006

26. Schmidt WA, Kraft HE, Völker L, Vorpahl K, Gromnica-Ihle EJ. Colour doppler sonography to diagnose temporal arteritis. The Lancet. 1995.

27. Aschwanden M, Daikeler T, Kesten F, Baldi T, Benz D, Tyndall A, et al. Temporal artery compression sign - A novel ultrasound finding for the diagnosis of giant cell arteritis. Ultraschall der Medizin. 2013;

28. Chrysidis S, Duftner C, Dejaco C, Schäfer VS, Ramiro S, Carrara G, et al. Definitions and reliability assessment of elementary ultrasound lesions in giant cell arteritis: A study from the OMERACT large vessel vasculitis ultrasound working group. RMD Open. 2018;

29. Touboul PJ, Hennerici MG, Meairs S, Adams H, Amarenco P, Bornstein N, et al. Mannheim carotid intima-media thickness and plaque consensus (2004-2006-2011). Cerebrovascular Diseases. 2012.

30. Skilton MR, Boussel L, Bonnet F, Bernard S, Douek PC, Moulin P, et al. Carotid intima-media and adventitial thickening: Comparison of new and established ultrasound and magnetic resonance imaging techniques. Atherosclerosis. 2011;

31. Schafer VS, Juche A, Ramiro S, Krause A. Cut-off values for intima-media thickness of temporal, facial and axillary arteries in giant cell arteritis. Ann Rheum Dis. 2016;

32. De Miguel E, Roxo A, Castillo C, Peiteado D, Villalba A, Martín-Mola E. The utility and sensitivity of colour Doppler ultrasound in monitoring changes in giant cell arteritis. Clin Exp Rheumatol. 2012;

33. Diamantopoulos AP, Haugeberg G, Hetland H, Soldal DM, Bie R, Myklebust G.

Diagnostic value of color doppler ultrasonography of temporal arteries and large vessels in giant cell arteritis: A consecutive case series. Arthritis Care Res. 2014;

34. Nagaraj Y, Asha CS, Teja A. HS, Narasimhadhan A V. Carotid wall segmentation in longitudinal ultrasound images using structured random forest. Comput Electr Eng. 2018;

35. Delsanto S, Molinari F, Giustetto P, Liboni W, Badalamenti S, Suri JS. Characterization of a completely user-independent algorithm for carotid artery segmentation in 2-D ultrasound images. IEEE Trans Instrum Meas. 2007;

36. Zhou Y, Cheng X, Xu X, Song E. Dynamic programming in parallel boundary detection with application to ultrasound intima-media segmentation. Med Image Anal. 2013;

37. Loizou CP, Pattichis CS, Pantziaris M, Tyllis T, Nicolaides A. Snakes based segmentation of the common carotid artery intima media. Med Biol Eng Comput. 2007;

38. Menchón-Lara RM, Sancho-Gómez JL. Fully automatic segmentation of ultrasound common carotid artery images based on machine learning. Neurocomputing. 2015;

39. Rothwell PM. External validity of randomised controlled trials: "To whom do the results of this trial apply?" Lancet. 2005;

40. Centre for clinical research and biostatistics (CCRB). C.I. Calculator: Diagnostic Statistics [Internet]. [cited 2018 Jan 12]. Available from: https://www2.ccrb.cuhk.edu.hk/stat/confidence interval/Diagnostic Statistic.htm#top

41. Martin Bland J, Altman DG. STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT. Lancet. 1986;

42. Hilgers R-D, Heussen N, Stanzel S. Bland-Altman-Plot. In: Lexikon der Medizinischen Laboratoriumsdiagnostik. 2018.

43. Sedgwick P. Limits of agreement (Bland-Altman method). BMJ (Online). 2013.

44. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;

45. Canny J. A Computational Approach to Edge Detection. IEEE Trans Pattern Anal Mach Intell. 1986;

46. Polesel A, Ramponi G, Mathews VJ. Image enhancement via adaptive unsharp masking. IEEE Trans Image Process. 2000;

47. Otsu N. A Threshold Selection Method from Gray-Level Histograms. IEEE Trans Syst Man Cybern. 2008;

48. The MathWorks. fspecial [Internet]. [cited 2018 Aug 11]. Available from: https://nl.mathworks.com/help/images/ref/fspecial.html

49. Altman DG, Bland j. M. Statistics Notes: Diagnostic tests 2: Predictive values. BMJ. 1994;

50. Laskou F, Coath F, Mackie SL, Banerjee S, Aung T, Dasgupta B. A probability score to aid the diagnosis of suspected giant cell arteritis. Clin Exp Rheumatol. 2019;